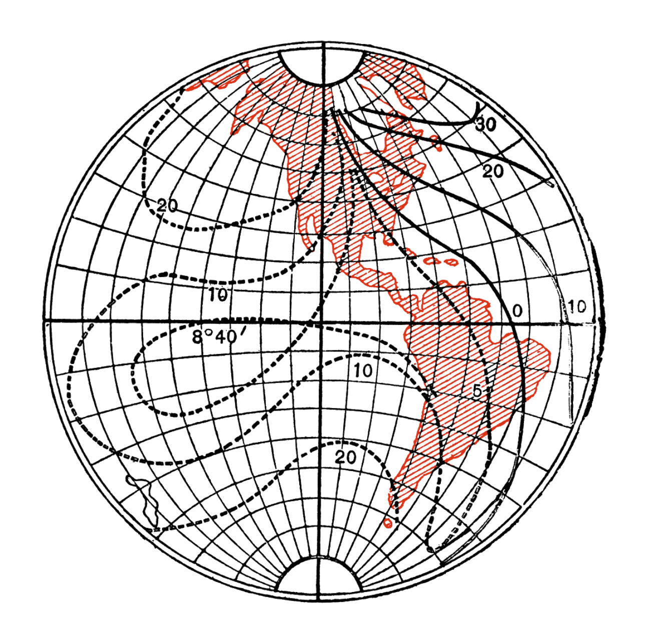


Sistemas de Información Geográfica



Sistemas de Información Geográfica

Víctor Olaya

2020

Sistemas de Información Geográfica

ISBN: 978-1-71677-766-0

Imagen de cubierta: Elroy M. Avery *School Physics* (New York: Sheldon and Company, 1895)

Versión revisada el 8 de julio de 2020

Introducción

Acerca de este libro

A principio de marzo de 2006, tuvieron lugar en Girona las I Jornadas sobre Sistemas de Información Geográfica (SIG) libres, organizadas por la Universitat de Girona y el Servei de Sistemes d'Informació Geogràfica i Teledetecció (SIGTE). Al amparo de ese encuentro, algunos (no muchos) decidimos aprovechar el momento para fundar el capítulo hispano–hablante de OSGeo, una organización internacional cuya actividad se centra en torno a los Sistemas de Información Geográfica libres. El objetivo era intentar formar un grupo de trabajo para afrontar aquellas cuestiones de interés general para la comunidad SIG de habla hispana, y fueron muchas las tareas que en aquel momento se recopilaron como necesarias dentro del ámbito de los SIG en lengua española. Entre ellas, la creación de un libro libre sobre Sistemas de Información Geográfica.

Este libro es mi intento particular de tratar de dar solución a esa necesidad, y recopila el trabajo realizado desde entonces, siendo un testigo de todo el esfuerzo y desarrollo que tiene lugar en nuestros días en materia de SIG. Con él, espero que dicho desarrollo no solo continúe, sino que crezca, poniendo a disposición de la comunidad los conocimientos con los que establecer una base conceptual más sólida, así como una herramienta para la formación y transmisión de los aspectos teóricos relacionados con los Sistemas de Información Geográfica.

¿Por qué escribir este libro?

Como ya quedó claro en aquella reunión fundacional del capítulo hispano–hablante de OSGeo, un libro como este era algo necesario dentro del mundo de los SIG, y lo sigue siendo hoy a pesar del tiempo que ha pasado desde entonces.

Es difícil encontrar hoy en día una disciplina científica que no pueda sacar partido de los Sistemas de Información Geográfica y no contemple a estos como herramientas de primera línea. Incluso fuera del ámbito científico, los SIG son parte de nuestra vida diaria, y la mayoría de la gente ha usado en algún momento un callejero en Internet o un navegador GPS, elementos que forman parte del conjunto de tecnologías directamente relacionadas con los SIG y su entorno.

En una situación así, cabe esperar que el sector editorial se encuentre en un estado de similar desarrollo, y el número de obras disponibles sea no solo abundante, sino diverso, y que cubra con detalle tanto los fundamentos básicos de la disciplina como los desarrollos más recientes (que no son pocos). Por desgracia, no sucede así. Si hablamos del mercado editorial en español, las referencias de que disponemos no están en absoluto acordes con el buen estado del ámbito de los SIG, tanto en su desarrollo e implementación como en el volumen de negocio que generan en los países de habla hispana.

Los principales textos de SIG en nuestro idioma se publicaron en los años 90, al tiempo que comenzaba a asentarse el sector de los Sistemas de Información Geográfica, especialmente en Es-

paña, ocupando un nicho entonces desierto dentro del mercado editorial. Lamentablemente, ninguno de estos trabajos se ha actualizado recientemente, a pesar de que los cambios que se han producido en el sector han sido tan notables que dicha actualización resulta completamente imprescindible.

A partir del año 2000, el ritmo editorial en el ámbito del SIG no decae, pero las referencias nuevas que hacen su aparición en el mercado son casi en su totalidad de corte práctico, bien sea referidas a un software en concreto, o bien a un campo de aplicación particular. Los fundamentos básicos, necesarios para cualquier operación con un SIG, quedan en estos casos explicados en base a un programa determinado, o bien se supone su conocimiento a través de otros textos.

En resumen, es un buen momento para escribir un libro sobre SIG en español, y es así porque resulta necesario recoger los conocimientos actuales en este campo, de la misma forma que se recogieron en una etapa anterior en algunos de esos textos comentados.

Pero la razón para escribir este libro no es únicamente que este sea un buen momento para escribir un libro de SIG. Porque este no es solo un libro de SIG, sino un libro *libre* sobre SIG. Y si este resulta un momento ideal para escribir un libro sobre SIG en nuestro idioma, lo es más aún si este libro es libre, tal y como el que ahora mismo estás leyendo. A pesar de la madurez de la disciplina de los SIG, no existen, ni en nuestro idioma ni en otros, referencias libres que detallen los fundamentos de este campo. Los textos libres que se encuentran disponibles van de la mano de aplicaciones (normalmente también libres), y constituyen manuales de uso o introducciones al mundo SIG apoyadas siempre en un software determinado. Encontrar una obra libre que se ocupe de tratar los fundamentos del SIG es, a día de hoy, tarea imposible.

Así pues, es un buen momento para escribir un libro en español sobre SIG, y es un excelente momento para hacer un libro libre sobre SIG, con independencia del idioma, pues ambos son elementos necesarios con los que no contamos. Hay muchos lectores potenciales que quieren aprender sobre SIG, y poner a su alcance los conocimientos actuales es la razón por la que he escrito el libro que ahora tienes en tus manos (o en la pantalla de tu ordenador o en cualquier otro soporte en el que, gracias a la libertad con la que se distribuye, puede haber llegado hasta ti).

A quién va dirigido este libro

Es importante reseñar que este es un libro principalmente teórico. Está enfocado a todos aquellos que deseen aprender los fundamentos teóricos de la disciplina de los SIG, abarcando la practica totalidad de ramas de esta. Estas ramas toman elementos de muchas otras disciplinas distintas (la informática, la geografía, la matemática...), y todas ellas se han tratado a su vez con suficiente nivel de detalle. Un lector que asimile la mayor parte de los conocimientos de este libro estará en una situación excepcional para comprender en conjunto todo lo que representan los SIG, qué tareas pueden hacerse con ellos, y entender por qué, cómo y cuándo se han de llevar a cabo dichas tareas. El objetivo es formar a alguien en materia de SIG, de tal modo que pueda posteriormente afrontar tareas relacionadas, tales como la realización de proyectos con Sistemas de Información Geográfica, o la aplicación de los SIG a un área concreta de trabajo, cualesquiera que sean las características de esta.

Aunque se trate de un libro teórico, este libro puede (y debe) convertirte en un buen usuario de SIG y ayudarte en el terreno práctico de su utilización. Tanto si eres ya usuario de SIG como si no lo eres, es probable que la lectura de estas

páginas te aporte gran cantidad de información que mejorará la forma en que utilizas (o utilizarás) un Sistema de Información Geográfica, pues te ayudará a entenderlo mejor. No obstante, si lo que pretendes es aprender a utilizar un SIG y buscas un manual para ello, este no es, en principio, el libro que andabas buscando, ya que no se dan indicaciones sobre cómo usar un SIG en concreto o cómo realizar procesos y operaciones con él. Podrías leer completo este libro y después no saber qué hacer ante una aplicación SIG, porque no vas a encontrar aquí cómo utilizar ninguna de ellas.

En ese caso, y sí deseas aprender a manejar un programa concreto pero no tienes experiencia previa en este campo, este libro será de todas formas una herramienta muy útil. Puedes complementar el contenido de estas páginas con el manual de usuario del SIG que elijas, e ir aprendiendo a utilizar este a medida que construyes una base sólida de conocimientos sobre la que apoyar dicho aprendizaje.

Así pues, este es un libro orientado a quienes desean aprender con mayor o menor profundidad qué es y para que sirve un SIG, aportándoles la base necesaria para que posteriormente puedan afrontar su trabajo con un SIG, tanto si ya tienen nociones al respecto como si todavía no han utilizado uno antes.

Si al leer este libro echas en falta algo que consideras importante, házmelo saber enviándome un mensaje a `volayaf@gmail.com`.

Agradecimientos

Las siguientes personas han colaborado en la creación de este libro: Juan Pedro Andersen, Ramiro Aznar, Landon Blake, Javier Carrasco, Amancio Delgado, Juan Carlos Giménez, Tomislav Hengl, Toni Hernández, Santiago Higuera, José Manuel Llorente, Ester López, Miguel Luaces, Jordi Marturià, Miguel Montesinos, Bernabé Moreno Ramírez, Rosa Olivella, Ferrán Orduña, Joana Palahí, Nuria Pérez, Carol Puig, Jorge Sanz, Josep Sitjar, David Tabernero, Ian Turton, Nacho Varela, Ana Velasco, Laura Vergoñós y Lluis Vicens. Mi más sincero agradecimiento a todos ellos por hacer posible este proyecto.

Índice general

Parte I

Los fundamentos

¿Qué es un SIG?

1.1. Introducción

La mayor parte de la de la información que manejamos en cualquier tipo de disciplina está georreferenciada. Es decir, se trata de información a la cual puede asignarse una posición geográfica, y es por tanto información que viene acompañada de otra información adicional relativa a su localización.

Si bien esto no es un hecho novedoso, la situación actual es más favorable que nunca para el desarrollo de herramientas que permitan el aprovechamiento de toda esa información, al tiempo que se consideran los datos relativos a su posición en el espacio. Hoy en día no solo trabajamos con gran cantidad de información referenciada geográficamente, sino que somos cada vez más conscientes de la importancia que esa componente geográfica tiene.

En ese contexto, un SIG es, fundamentalmente, una herramienta para trabajar con información georreferenciada. Dentro de esta definición entran un gran número de tecnologías y de otros elementos no tecnológicos, los cuales veremos a lo largo de este libro.

1.2. Un pequeño ejemplo

Para comenzar a tener una idea correcta de lo que representa e implica un SIG, veamos un sencillo ejemplo. Supongamos el caso de un organismo o empresa cuyo trabajo incluye la gestión de una masa forestal. Este trabajo de gestión implicará algunas actividades como las siguientes, en las cuales se utiliza en mayor o menor medida información georreferenciada.

- Delimitación de las distintas zonas inventariables y unidades de gestión (montes, cantones, rodales, etc.)
- Diseño de inventarios
- Realización de inventarios y gestión de sus datos para la obtención de resultados tales como estimaciones de volúmenes maderables.
- Gestión de infraestructuras del monte tales como vías de comunicación, torres de vigilancia contra incendios, etc.

En un contexto en el que no existen medios informáticos para la realización de estas tareas, gran parte de ellas se desarrollarán con el apoyo de cartografía clásica. Así, las zonas inventariables se delimitarán sobre un plano, y sobre este mismo pueden medirse sus superficies con la ayuda de un planímetro. En ese mismo plano se localizan las parcelas a muestrear en un inventario, y los operarios encargados de llegar hasta esas parcelas y realizar las mediciones pertinentes se ayudan de él para localizarlas y desplazarse sobre el terreno.

Los resultados del inventario se almacenan en estadillos, y las operaciones correspondientes al análisis estadístico de estos se realizan de forma manual, así como la comparación con inventarios

anteriores que permiten estudiar la evolución del monte.

La presencia de medios informáticos facilita estas tareas, mejorando por una parte la gestión de los datos, y por otra las operaciones que pueden realizarse sobre estos. Una sencilla hoja de cálculo, por ejemplo, es una herramienta imprescindible para la gestión de los datos de un inventario, haciendo que todo el trabajo con ellos resulte más eficiente y adecuado.

En lo relativo a la cartografía, la situación, aunque con un desarrollo (y especialmente una implantación de usuarios) más reciente, no es muy distinta. Ventajas similares a las que aporta una hoja de cálculo pueden encontrarse en una aplicación que permitiera utilizar mapas y planos *dentro* de un ordenador, con la consecuente ganancia en productividad, eficiencia y precisión. Esta aplicación destinada al manejo de cartografía es el concepto básico de un Sistema de Información Geográfica, y la idea fundamental a partir de la cual comenzó el desarrollo de estos.

Con un SIG, la cartografía de esa masa forestal puede visualizarse y almacenarse en un ordenador personal, y pueden realizarse sin dificultad y de forma instantánea cálculos tales como mediciones de cada una de las entidades. La creación de nueva información cartográfica se lleva a cabo en el propio SIG, del mismo modo que la edición de cartografía ya existente. Modificar el límite de una unidad de gestión o el trazado de una vía, o crear la cartografía correspondiente a las parcelas de inventario, son tareas que, en nuestro caso de ejemplo, se realizan hoy en día empleando un SIG.

Las ventajas que esto tiene son muchas, especialmente las relacionadas con una mejor gestión del conjunto de distintos datos que se manejan, así como las relativas a la sencillez con que pueden modificarse estos datos[1].

Otras de las labores donde un SIG demuestra su utilidad es en el análisis. Los datos geográficos pueden ser objeto de gran número de distintos análisis, y la capacidad de cómputo de un ordenador es necesaria para muchos de ellos. La herramienta idónea para implementar esos algoritmos y operaciones de análisis espacial es el SIG, pues ya contiene los elementos necesarios para el manejo de los datos de partida, es decir, aquellos que contienen la información georreferenciada.

Y, por supuesto, un SIG conectado a un periférico de impresión permite generar una versión analógica a partir de la información con la que se trabaja, teniendo la capacidad de crear cartografía en papel cuando así se requiera.

En otras palabras, un SIG es una herramienta que brinda a las labores de uso y manejo de información geográfica toda la potencia de un ordenador, pues ha sido diseñada específicamente para trabajar con este tipo particular de información.

No obstante, más allá de todas estas tareas antes mencionadas, el concepto de SIG ha evolucionado hasta convertir actualmente a estos en sistemas complejos que buscan dar solución a todas las necesidades que se presentan en situaciones similares a la del ejemplo anterior. Con la tecnología actual, la incorporación de elementos propios de los SIG puede llegar mucho más allá, y uno de los pilares más sólidos de los SIG en la actualidad es su capacidad de mostrar que existe una componente espacial susceptible de ser gestionada con la ayuda de un SIG en la práctica totalidad de contextos posibles.

Como sistema, un SIG puede gestionar la cartografía necesaria para la gestión integral del monte, y hacerlo además de forma centralizada. De este modo, se garantiza el rigor y la robustez de los datos base, ya que el SIG es el encargado de canalizar la utilización de estos por parte de

[1]Veremos con más detalle las ventajas de los datos digitales frente a los datos analógicos en el capítulo 6

todos los usuarios. Esto es de especial importancia en caso de que se editen los datos, ya que esta edición también está centralizada, y un usuario ve reflejarse en su cartografía de forma inmediata los cambios realizados por otro, teniendo siempre a su disposición la versión más actual y, por tanto, más adecuada.

A esto puede añadirse la utilización de SIG móviles en dispositivos portátiles, que permiten que el SIG se incorpore también a las fases de trabajo de campo. Esa misma cartografía centralizada pueden utilizarla los operarios en campo a través de sus dispositivos para desarrollar su trabajo, ayudándose además de sistemas de navegación para la localización de las parcelas de un muestreo o de cualquier otro punto de interés al que deban desplazarse.

Gracias a la tecnología SIG, la información espacial puede ser aprovechada en mayor medida, y en muchos casos pasa de ser una información inherente a los datos pero sin una verdadera aplicación, a ser un elemento sumamente enriquecedor y clave para muchos análisis.

En nuestro ejemplo de gestión forestal, los propios datos del inventario, que antes eran fundamentalmente datos sobre las propiedades de los distintos árboles medidos (altura, diámetro, etc.), ahora ofrecen muchas más posibilidades si se considera que cada uno de estos árboles ha sido medido en una parcela dada, la cual lleva asociadas unas coordenadas concretas.

El trabajo que se desarrollaba en la hoja de cálculo con estos datos se puede incorporar al SIG, el cual además de las funciones de análisis estadístico incluye funciones de análisis espacial. De este modo, los resultados numéricos que se obtenían de esos análisis (volúmenes totales estimados, alturas medias, etc.) se amplían mediante resultados con mayor componente espacial, como puede ser la creación de nueva cartografía referente a las variables principales (mapas de densidad media de arbolado, altura dominante media, etc.).

En resumen, el SIG en su concepción actual es una herramienta integradora que busca abarcar en su ámbito todas las funcionalidades que se requieren para el trabajo con variables y elementos espacialmente localizados, incorporando para ello capacidades variadas que serán las que vayamos viendo progresivamente a lo largo de esta obra.

1.3. ¿Qué es un SIG?

Partiendo del ejemplo anterior, podemos dar una definición más precisa y formal de lo que realmente es un SIG. Básicamente, un SIG ha de permitir la realización las siguientes operaciones:

- Lectura, edición, almacenamiento y, en términos generales, gestión de datos espaciales.
- Análisis de dichos datos. Esto puede incluir desde consultas sencillas a la elaboración de complejos modelos, y puede llevarse a cabo tanto sobre la componente espacial de los datos (la localización de cada valor o elemento) como sobre la componente temática (el valor o el elemento en sí).
- Generación de resultados tales como mapas, informes, gráficos, etc.

En función de cuál de estos aspectos se valore como más importante, encontramos distintas definiciones formales del concepto de un SIG. Una definición clásica es la de [1], para quien un SIG es un elemento que permite «analizar, presentar e interpretar hechos relativos a la superficie terrestre». El mismo autor argumenta, no obstante, que «esta es una definición muy amplia, y habitualmente se emplea otra más concreta. En palabras habituales, un SIG es un conjunto de *software* y *hardware* diseñado específicamente para la adquisición, mantenimiento y uso de datos cartográficos».

En una línea similar, [2] define un SIG como un «sistema de información diseñado para trabajar con datos referenciados mediante coordenadas espaciales o geográficas. En otras palabras, un SIG es tanto un sistema de base de datos con capacidades específicas para datos georreferenciados, como un conjunto de operaciones para trabajar con esos datos. En cierto modo, un SIG es un mapa de orden superior».

Ambas definiciones recogen el concepto fundamental de los SIG en el momento en que fueron escritas, pero la realidad hoy en día hace necesario recoger otras ideas, y la definición actual de un SIG debe fundamentarse sobre todo en el concepto de *sistema* como elemento integrador que engloba a un conjunto de componentes interrelacionados.

Como apunta [1], *software* y *hardware* son dos elementos primordiales del SIG, pero no son sin embargo los únicos. En el contexto actual, otros componentes juegan un papel igual de importante en la idea global de un SIG.

De igual modo, un SIG puede considerarse como un «mapa de orden superior» entendiendo que se trata de una forma más potente y avanzada de hacer todo aquello que, previamente a la aparición de los SIG, se llevaba a cabo mediante el uso de mapas y cartografía en sentido clásico. Es decir, los SIG representan un paso más allá de los mapas. No obstante, esta definición resulta en exceso simplista, pues mapas y SIG no son conceptos equiparables en el contexto actual de estos últimos.

Un mapa es una representación de un conjunto de datos espaciales y, aunque esta representación resulta de enorme importancia, en el entorno de un SIG no es sino un elemento más de una serie de componentes (tales como el *software* y el *hardware* que antes mencionábamos). Más aún, un SIG contiene no solo los datos y la representación, sino también las operaciones que pueden hacerse sobre el mapa, que no son ajenas a este sino partes igualmente de todo el sistema conformado por el SIG.

De la misma forma que los textos han pasado del papel al ordenador (antes leíamos libros, ahora podemos leer libros impresos, libros digitales, páginas Web, etc.), los mapas también han dado ese salto cualitativo con la aparición de los SIG. Sin embargo, el SIG es mucho más que una nueva forma de cartografía, y no invalida en absoluto formas anteriores. De hecho, una función muy importante de los SIG es ayudar a crear mapas en papel, y estos se siguen utilizando hoy en día en todos los ámbitos. Y junto con esta funcionalidad, encontramos otras que hacen que en su conjunto un SIG sea una herramienta integradora y completa para el trabajo con información georreferenciada.

Debe entenderse, pues, un SIG como un elemento complejo que engloba una serie de otros elementos conectados, cada uno de los cuales desempeña una función particular. Estos elementos son, como iremos viendo más adelante, los datos, los procesos, la visualización, la tecnología y el factor organizativo. Baste por el momento citarlos, ya que más adelante, y a lo largo de todo el libro, se irán describiendo pormenorizadamente todos ellos.

Con lo anterior, una definición más precisa es decir que un SIG es un sistema que integra tecnología informática, personas e información geográfica, y cuya principal función es capturar, analizar, almacenar, editar y representar datos georreferenciados [3].

1.3.1. SIG como integrador de información

Si bien un SIG tiene una inherente naturaleza integradora y esta puede enfocarse desde muchos puntos de vista, el elemento más relevan-

te es la propia información que un SIG maneja y las características de esta. Conceptualmente, el verdadero pilar de esa naturaleza integradora del SIG reside en la información geográfica con la que se trabaja, que provee la amalgama adecuada para que un SIG sea un sistema sólido y cohesionado.

Muchas disciplinas trabajan con información de distinta naturaleza. En ellas, no siempre resulta sencillo buscar elementos en común para poder unir y coordinar toda esa información bajo un único punto de vista conceptual. En otras ocasiones, disciplinas que en la práctica presentan una interacción real (puede decirse que, de un modo u otro, todas las disciplinas están interrelacionadas) resultan difíciles de integrar desde el punto de vista teórico, y no es sencillo ponerlas en un marco común de trabajo.

Por ejemplo, información de tipo sociológico como la tasa de analfabetismo e información de carácter físico o biológico como puede ser la acidez del suelo, no parecen sencillas de combinar para la realización de algún análisis común. De existir alguna relación entre ellas (o de no existir, y pretender demostrar que son variables independientes), es necesario buscar un punto de enlace entre ambas informaciones para poder estudiar esta. Un nexo que las une es el hecho de que están asociadas a una localización en el espacio, ya que una serie de datos de tasa de analfabetismo corresponderán a una serie de lugares, del mismo modo que lo harán los valores de acidez del suelo.

El hecho de que ambas informaciones tienen a su vez carácter geográfico va a permitir combinarlas y obtener resultados a partir de un análisis común. Puesto que una gran parte de la información está georreferenciada, esa georreferencia va a representar un punto común para enmarcar el análisis. El SIG es, en este contexto, el marco necesario en el que incorporar esa información georreferenciada y trabajar con ella.

1.3.2. SIG como integrador de tecnologías

Puede pensarse que los SIG son meramente herramientas informáticas y que la única tecnología que reside tras ellas es la propia tecnología informática. Sin embargo, el papel integrador de los SIG hace que sean la herramienta elegida para la gestión de resultados y elementos producidos por otras tecnologías, muchas de las cuales se encuentran actualmente en pleno desarrollo.

La popularización de los SIG y su mayor presencia en una buena parte de los ámbitos de trabajo actuales han traído como consecuencia una mayor conciencia acerca de la importancia de la componente espacial de la información, así como sobre las posibilidades que la utilización de esta ofrece. Por ello, una gran parte de las tecnologías que han surgido en los últimos años (y seguramente de las que surjan en los próximos) se centran en el aprovechamiento de la información espacial, y están conectadas en mayor o menor medida a un SIG para ampliar su alcance y sus capacidades. Por su posición central en el conjunto de todas las tecnologías, los SIG cumplen además un papel de unión entre ellas, conectándolas y permitiendo una relación fluida alrededor de las funcionalidades y elementos base de un Sistema de Información Geográfica.

1.3.3. SIG como integrador de personas

Ya sabemos que la información georrefenciada es muy numerosa y variada. Esto significa que son muchos los tipos de personas que pueden emplearla y, por tanto, que pueden emplear un SIG para el trabajo con ella. La presencia del SIG como puerta de acceso a esa información es un punto común a todas esas distintas personas, y un Sistema de Información Geográfica es tam-

bién un elemento integrador a nivel humano y profesional.

Dentro incluso de un mismo campo de aplicación, son varios los grupos de personas que van a estar implicados en el desarrollo de una tarea dada con la ayuda de un SIG. Desde la creación del dato geográfico hasta la obtención de un resultado final, son muchas las operaciones que se llevan a cabo, y estas las desarrollan profesionales de distinta especialización y con herramientas adaptadas a dichas operaciones. En nuestro ejemplo, y en la etapa previa a la aparición de los SIG, las herramientas que empleaba el cartógrafo para generar un mapa eran diferentes de las que empleaba el gestor para analizar dicho mapa, y estas a su vez distintas a las que podían emplearse para la elaboración de resultados.

Con la aparición de los SIG, todos los profesionales dentro de esa cadena que va desde la creación del dato hasta las operaciones finales que se realizan sobre estos tienen una herramienta común de trabajo, pues un SIG puede utilizarse para desarrollar parcial o totalmente las tareas correspondientes a cada uno de ellos. El SIG es empleado para crear cartografía, para almacenar, gestionar y consultar esta, así como para realizar análisis más complejos en base a ella y crear resultados.

Las funciones básicas que un SIG ha de cumplir, que ya vimos en el momento de dar una definición de estos, cubren en realidad un rango amplio de trabajo, y engloban las necesidades de usuarios que con anterioridad no tenían entre sí un marco de trabajo común tan definido. Esto tiene como consecuencia que existe una mejor coordinación entre ellos, pues es la propia herramienta quien establece las características de la relaciones existentes, y estas no dependen ya únicamente del propio ámbito de aplicación.

1.3.4. SIG como integrador de teorías y fundamentos. La Ciencia de la Información Geográfica

La evolución conceptual que se ha producido en el ámbito de los SIG, pasando como ya hemos visto de ser considerados simples programas informáticos a sistemas completos con múltiples componentes, ha tenido lugar también en la ciencia que los rodea. Los SIG no solo han contribuido al desarrollo de las ciencias afines, sino que en muchos casos han modificado estas o han contribuido a la formación de nuevas ramas. Conceptos básicos y hasta ese momento sólidos, como por ejemplo la idea de lo que es y lo que significa un mapa (una idea fundamental para el trabajo en muchas disciplinas), han sido literalmente redefinidas desde la aparición de los SIG.

Desde un punto de vista muy simple, podemos entender un SIG como la unión de dos ciencias: la geografía y la informática. Visto así, un SIG es una herramienta informática para ayudar al trabajo en el ámbito geográfico. Esta concepción tan simple dista, no obstante, mucho del concepto real de un SIG, pues este incorpora elementos de muchas ciencias distintas como pueden ser las siguientes[4]:

- Disciplinas relacionadas con la tecnología y el manejo de información. Se incluyen aquí las ciencias de la información, la informática, el diseño de bases de datos o el tratamiento digital de imágenes, entre otras. Muchas de estas, a su vez, derivan de otras o toman importantes elementos de ellas. La estadística o la matemática son algunas de esas ciencias fundamentales.
- Disciplinas dedicadas al estudio de la Tierra desde un punto de vista físico. La geología, la oceanografía, la ecología, así como todo el conjunto de ciencias medioambientales, forman parte de este grupo.

- Disciplinas dedicadas al estudio de la Tierra desde un punto de vista social y humano. En este grupo se incluyen la antropología, la geografía o la sociología, entre otras. Las ciencias de este grupo, así como las del anterior, son todas ellas potenciales usuarias de los SIG.
- Disciplinas dedicadas al estudio del entendimiento humano, en particular en lo concerniente a la interacción con máquinas. Las ciencias del conocimiento, la psicología en general o las ramas que estudian y desarrollan la Inteligencia Artificial también juegan su papel en el contexto actual de los SIG.
- Disciplinas que tradicionalmente han realizado una integración de conocimientos de otros ámbitos distintos. La geografía es la principal representante de este grupo.

En el contexto presente, podemos entender la Ciencia de la Información Geográfica[2] como todo el conjunto de disciplinas y conocimientos que residen tras los SIG, tanto en su desarrollo y creación como en su utilización y aspectos prácticos. Esta ciencia se enmarcaría a su vez dentro de ese último grupo de disciplinas integradoras, llevando más allá la idea de la geografía como área de conocimiento que engloba elementos de muchos otros ámbitos.

El término *geomática*, formado a partir de los vocablos *geografía* e *informática*, se emplea con frecuencia para hacer mención a todo ese grupo de ciencias relacionadas con los SIG. No obstante, no se refiere exclusivamente a esas dos disciplinas, sino que simplemente toma nombre de los dos bloques principales de conocimiento a partir de los cuales se ha desarrollado la ciencia de los SIG.

Si los SIG deben ser entendidos a día de hoy como un sistema, la ciencia que los define y en la que se fundamentan debe no solo describir y servir de soporte a su elementos, sino también atender a una de las características fundamentales de todo sistema: las interrelaciones existentes entre dichos elementos. Por esta razón, disciplinas tales como las ciencias del conocimiento juegan un papel importante en el ámbito de los SIG, pues son fundamentales para estudiar las relaciones entre sus componentes.

En este libro desarrollaremos elementos provenientes de distintas disciplinas, centrándonos en aquellas ramas que tengan mayor relevancia desde el punto de vista del usuario de SIG, y con independencia de cuál sea la funcionalidad que este pueda buscar. Dejaremos de lado algunos aspectos sin duda importantes pero que atañen a otros enfoques distintos, como pueden ser, por ejemplo, el desarrollo de aplicaciones SIG o el diseño de entornos SIG corporativos.

1.4. ¿Qué no es un SIG?

Es obvio que, pese a que su propia denominación indica específicamente que los SIG desarrollan su actividad con información geográfica y esta es necesaria para el trabajo con ellos, existen otras tecnologías que también pueden hacer uso directo de esa información y explotarla de formas alternativas. A medida que se ha ido redefiniendo el concepto de SIG, muchos elementos han ido entrando en el amplio paraguas actual del SIG, así como distintas disciplinas, según hemos visto y veremos más adelante. No obstante, esas propias disciplinas no han desaparecido como tales, y siguen existiendo de forma autónoma. Y cada una de ellas dispone de sus propias herramientas, las cuales pueden incluir también tecnologías o sistemas más complejos similares a los SIG pero con un enfoque distinto.

Dos son las principales soluciones que deben conocerse por su relación directa con el ámbito

[2] *Geographic Information Science* en inglés, abreviado como GIScience o simplemente con el propio acrónimo GIS

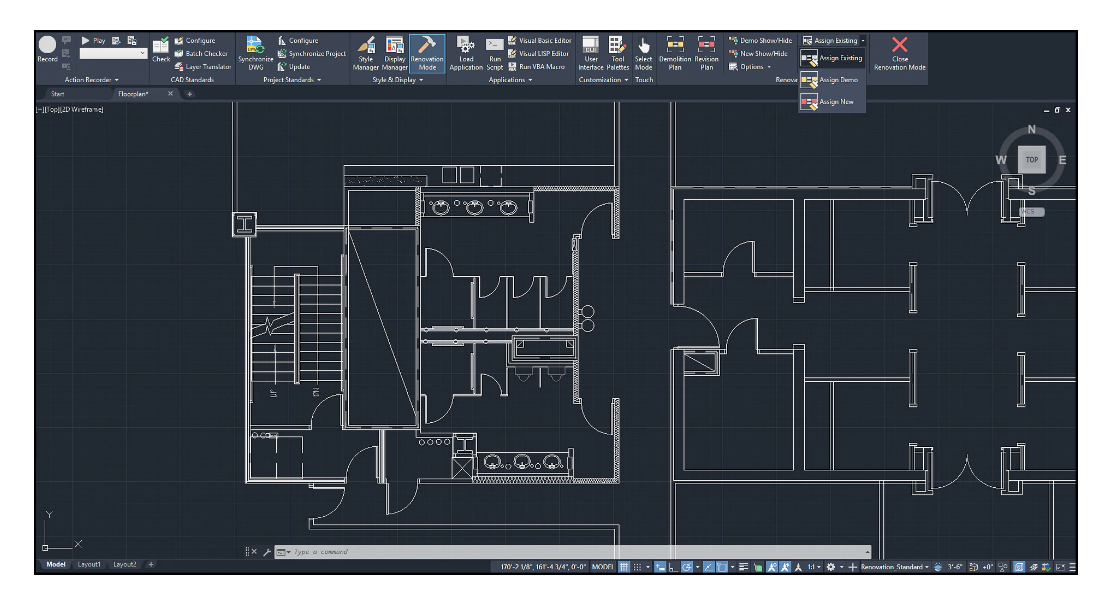


Figura 1.1: Aspecto de una aplicación CAD.

SIG: Diseño Asistido por Ordenador (CAD[3]) y AM/FM.

Las aplicaciones CAD (Figura 1.1) permiten el diseño informatizado de elementos muy diversos, que pueden ir desde una pieza industrial o la carrocería de un automóvil (tareas con poca relación con los SIG) a un edificio (con mayor relación con los SIG). El uso de herramientas CAD en disciplinas como la arquitectura para la creación de planos tiene cierta similitud con el uso de un SIG, y ambas herramientas se han nutrido la una de la otra en cuanto a sus funcionalidades. No obstante, siguen existiendo grandes diferencias que hacen que cada aplicación responda a unas necesidades concretas pese a la existencia de características comunes. De entre estas diferencias cabe destacar las siguientes [5][6]

- SIG y CAD han sido diseñados para propósitos diferentes. El del SIG es reflejar la realidad, mientras que el del CAD es diseñar algo que no existe todavía. La creación es el elemento fundamental en el CAD, mientras que el estudio de una realidad ya creada constituye la base del SIG.
- El almacenamiento de datos es diferente debido al distinto enfoque. En los SIG se da mayor peso a la gestión de los datos, mientras que en el CAD la parte visual es preponderante, y el almacenamiento así lo refleja. Un dato SIG se almacena como un dato geográfico complejo, mientras que en un CAD se almacena básicamente como un «dibujo», pues es ese el enfoque fundamental de trabajo.
- El volumen de datos en un SIG es ordenes de magnitud mayor, y ello implica una gestión de datos distinta y unas necesidades más elevadas en ese sentido. La escala de trabajo también alcanza dimensiones mayores, ya que, mientras que con ambas herramientas puede trabajarse en una extensión limitada, un CAD no esta pensado para gestionar datos de una superficie como la de un país, un continente o el planeta entero.

[3]Computer–Aided Design

- No todos los tipos de datos de un SIG se pueden incorporar en un CAD. Los datos procedentes de la teledetección, por ejemplo, no forman parte del abanico de datos que un CAD puede manejar.

El CAD puede resultar suficiente para desarrollar algunas tareas propias de los SIG, en particular las relacionadas con el diseño cartográfico. Aun así, una herramienta CAD no puede sustituir completamente a un SIG, ya que no ofrece una respuesta a requerimientos tales como:

- Análisis, modelización, y gestión avanzada de datos espaciales.
- Trabajo con datos que cubren una gran superficie geográfica.
- Necesidad de utilizar diversos sistemas de proyección.
- Edición de datos por usuarios de distinto perfil y de modo concurrente.

Por su parte, las siglas AM/FM(*Automated Mapping/Facilities Management*)[4] de uso poco habitual en nuestro idioma, hacen referencia a aplicaciones diseñadas para la gestión de infraestructuras generalmente de carácter público, tales como redes de alcantarillado, conducciones de gas o vías de circulación, entre otras.

Las aplicaciones empleadas para estas tareas tienen dos bloques básicos: un bloque gráfico de visualización y otro de gestión de datos. Este último almacena los atributos asociados a los elementos gráficos, que son principalmente de tipo lineal (tuberías, redes de alumbrado, etc.). Otro tipo de elementos, tales como elementos poligonales, son difíciles de manejar en estos sistemas, ya que su diseño obedece a las necesidades existentes en su ámbito de utilización, y estas se sitúan mayoritariamente alrededor de las infraestructuras lineales.

Sin embargo, incluso con este tipo de elementos las capacidades de una aplicación AM/FM no igualan a las de un SIG, ya que no incorporan otro tipo de información como la relativa a la topología (que describiremos con detalle en el capítulo 5). Esto es así debido a que el subsistema de análisis, fundamental en un SIG, no tiene presencia en estas herramientas, y por tanto sus características no incluyen aquellos componentes que sean necesarios exclusivamente para procesos de tipo analítico.

Puede decirse, por tanto, que este tipo de aplicaciones representa un subconjunto de los SIG, pues sus funcionalidades principales son más reducidas que las de estos, y su ámbito de aplicación es menos generalista. En cierta medida, las aplicaciones AM/FM se asemejan también a las aplicaciones CAD, poniendo un énfasis especial en la componente gráfica, aunque con una mayor adaptación a la naturaleza geográfica de la información con la que se trabaja.

Al contrario, sin embargo, de lo que sucede con las aplicaciones CAD, en la actualidad las labores propias asociadas a los productos AM/FM se pueden llevar a cabo en un SIG genérico, o bien en una adaptación de este que tenga en consideración las características particulares del ámbito de trabajo. En este sentido, la gestión de servicios no es una aplicación más específica que otras a la hora de emplear un SIG, y este en la actualidad engloba de forma casi completa las funcionalidades de una herramienta AM/FM.

1.5. Componentes de un SIG

Como ya hemos visto, en su concepción actual los SIG son sistemas complejos que integran una serie de distintos elementos interrelacionados. El estudio de todos y cada uno de estos elementos es el fundamento para el estudio global de los Sistemas de Información Geográfica, y de ese modo se aborda a lo largo de este libro, mostran-

[4] Cartografía Automatizada/Gestión de Servicios

do las propias características de cada elemento y los conceptos necesarios para entender las relaciones entre ellos.

Una forma de entender el sistema SIG es como formado por una serie de subsistemas, cada uno de ellos encargado de una serie de funciones particulares. Es habitual citar tres subsistemas fundamentales:

- Subsistema de **datos**. Se encarga de las operaciones de entrada y salida de datos, y la gestión de estos dentro del SIG. Permite a los otros subsistemas tener acceso a los datos y realizar sus funciones en base a ellos.
- Subsistema de **visualización y creación cartográfica**. Crea representaciones a partir de los datos (mapas, leyendas, etc.), permitiendo así la interacción con ellos. Entre otras, incorpora también las funcionalidades de edición.
- Subsistema de **análisis**. Contiene métodos y procesos para el análisis de los datos geográficos.

Para que un SIG pueda considerarse una herramienta útil y válida con carácter general, debe incorporar estos tres subsistemas en cierta medida[7].

Otra forma distinta de ver el sistema SIG es atendiendo a los elementos básicos que lo componen. Cinco son los elementos principales que se contemplan tradicionalmente en este aspecto:

- **Datos**. Los datos son la materia prima necesaria para el trabajo en un SIG, y los que contienen la información geográfica vital para la propia existencia de los SIG.
- **Métodos**. Un conjunto de formulaciones y metodologías a aplicar sobre los datos.
- **Software**. Es necesaria una aplicación informática que pueda trabajar con los datos e implemente los métodos anteriores.
- **Hardware**. El equipo necesario para ejecutar el software.
- **Personas**. Las personas son las encargadas de diseñar y utilizar el software, siendo el motor del sistema SIG.

Para el enfoque de esta obra, cada uno de los elementos anteriores tiene unas características propias que deben estudiarse. No obstante, el hardware no es un elemento especialmente particular en el caso de un SIG, y las aplicaciones SIG que encontramos actualmente en el mercado en todas sus variedades (que son las que el lector de este libro va a utilizar habitualmente) se ejecutan en su mayoría sobre ordenadores personales sin requerimientos altamente específicos. Más aún, la expansión de las tecnologías SIG ha alcanzado hoy en día otros ámbitos como las plataformas móviles, haciendo de estas unas tecnologías poco específicas en lo que a *hardware* se refiere. Por esta razón, no es necesario tratar en detalle esta pieza del sistema SIG, siendo más adecuado tratar el resto de elementos, más característicos e importantes para el aprendizaje de los conceptos SIG y la descripción de estos.

Por su parte, las personas tienen importancia tanto de forma individual como en su conjunto, siendo diferentes las necesidades que plantean como usuarios y beneficiarios de un SIG. En la sociedad actual, las tecnologías y planteamientos colaborativos han calado hondo en el ámbito SIG, y la información geográfica es, por su propia naturaleza, propensa a ser compartida y utilizada por diferentes personas con fines muy distintos. Es por ello que el aspecto de mayor relevancia respecto a las personas como partes del sistema SIG es el de sus relaciones y su organización, siendo además en este campo donde se han producido en mayor medida los últimos avances, y donde ha tenido lugar un cambio más profundo, no ya solo dentro de los SIG, sino también en otras tecnologías de similar índole.

Puede entenderse esto como un nuevo subsistema: el subsistema *de gestión*, que es responsable

de gestionar la interacción de los restantes y definir y controlar el marco en que esta tiene lugar.

Las personas a su vez dan forma a los distintos ámbitos de trabajo, definiendo estos en función de sus necesidades. Puede tratarse el conjunto de campos de especialización como un nuevo elemento del sistema SIG, en lugar de incorporarlo dentro de otro.

Algunos autores proponen modificar el esquema clásico de cinco elementos para reflejar más correctamente la nueva realidad de los SIG. Por ejemplo, es habitual la propuesta de un esquema en el que se incluye la visualización como elemento por derecho propio. Si volvemos al enfoque basado en subsistemas, el subsistema de visualización resulta de enorme importancia en un SIG, siendo pese a ello habitual que no se trate con la suficiente profundidad en textos dedicados a los SIG desde un punto de vista genérico. Precisamente por no ser considerado un elemento independiente, no se le concede la necesaria atención como parte que debe estudiarse al tratar la disciplina de los SIG.

Con todo lo anterior, resulta más conveniente para su estudio práctico adoptar una evolución del esquema clásico de cinco elementos, y establecer unos nuevos componentes, cada uno de los cuales actúa como un pilar conceptual sobre es que ha de sustentarse es estudio de la disciplina de los SIG. Estos componentes son cinco:

- **Datos**.
- **Análisis**. Métodos y procesos enfocados al análisis de los datos.
- **Visualización**. Métodos y fundamentos relacionados con la representación de los datos.
- **Tecnología**. *Software* y *hardware* SIG
- **Factor organizativo**. Engloba los elementos relativos a la coordinación entre personas, datos y tecnología, o la comunicación entre ellos, entre otros aspectos.

A modo de introducción, se describen a continuación algunas ideas básicas de cada uno de estos componentes. Salvo el factor organizativo, cuyos elementos se explicarán a lo largo de toda esta obra, el resto de componentes se detallarán individualmente, cada uno de ellos en una parte correspondiente del libro.

1.5.1. Datos

Los datos son necesarios para hacer que el resto de componentes de un SIG cobre sentido y puedan ejercer su papel en el sistema. La información geográfica, la verdadera razón de ser los SIG, reside en los datos, y es por ello que el conocimiento exhaustivo de los datos y su naturaleza resulta obligado para una buena comprensión de los propios SIG.

Son muchas las facetas de los datos que deben estudiarse, y todas ellas con una gran importancia. Por un lado, es necesario conocer las características fundamentales del dato geográfico que utilizamos en un SIG, es decir, su forma y sus propiedades. De ellas dependen, por ejemplo, los procesos que podremos o no realizar con los datos, y en general todo cuanto podemos esperar de ellos.

Prescindiendo del hecho de que se trata de un dato geográfico, es relevante conocer cómo los datos se gestionan y almacenan en un entorno digital, aspectos de corte puramente informático que desarrolla la disciplina de la gestión de bases de datos. Cuando las ideas fundamentales al respecto se aplican al caso particular de los datos geográficos, surgen conceptos que resultan básicos para un buen uso de un SIG, y que además van siendo cada vez más relevantes a medida que los volúmenes de datos de que se dispone van aumentando.

Al igual que aumenta el volumen de datos, lo hacen los orígenes de estos y las formas en que

la información geográfica puede recogerse. Un aspecto clave para una utilización correcta de un SIG es saber integrar datos de distinta procedencia, para lo cual es vital entender cómo esta afecta a las propias características de dichos datos.

Otros elementos tales como la calidad de los datos, la cual cobra cada día más importancia, serán tratados igualmente junto a los anteriores en una parte específicamente dedicada a los datos, probablemente una de las más importantes dentro de este libro.

1.5.2. Análisis

El análisis es una las funcionalidades básicas de los SIG, y una de las razones fundamentales que llevaron al desarrollo de estos. Un ordenador es una herramienta con enorme capacidad de cálculo, y esta puede aplicarse a los datos espaciales para obtener resultados de muy diversa índole.

En mayor o menor medida, un SIG siempre incorpora una serie de formulaciones que permiten la obtención de resultados y el análisis de los datos espaciales. Estas formulaciones representan procesos que pueden ser sumamente sencillos o enormemente complejos, y que pueden resultar de aplicación en uno u otro campo, o incluso con carácter general. Su origen puede ser muy variado, y no derivan necesariamente del ámbito puro de la geografía, sino que pueden ir desde simples consultas o mediciones a elaborados modelos que empleen datos de variables muy numerosas y arrojen resultados complejos. La estadística, entre otras ciencias, puede aportar al ámbito SIG muchas de sus ideas, y estas, adaptadas al marco de la información georreferenciada, constituir en el SIG un nuevo conjunto de procesos de análisis.

Las ventajas de la incorporación de todos estos procesos en una única herramienta, el SIG, van desde la automatización de tareas a la aparición de nuevos procesos que, aprovechando la gran capacidad de cómputo de la plataforma en la que se ejecuta el SIG, producen resultados que no podrían ser obtenidos de otro modo. Bien sea por la complejidad propia de los procesos o por el nivel de precisión al que se trabaja, existen muchos procesos que no pueden realizarse mediante el uso de cartografía clásica y sin el apoyo de medios informatizados. El SIG abre un campo de actuación en el que la práctica totalidad de ideas y formulaciones de análisis pueden plasmarse y aplicarse con carácter práctico.

1.5.3. Visualización

Cualquier tipo de información puede ser representada de forma gráfica, lo cual habitualmente facilita la interpretación de dicha información o parte de esta. Gran parte de las características de la información (por ejemplo, la presencia de patrones sistemáticos), son más fáciles de estudiar cuando se apoyan sobre algún elemento visual, pues este añade un nuevo punto de vista.

En el caso particular de la información geográfica, la visualización no solo es una forma más de trabajar con ella, sino que resulta la forma principal, no ya por ser la que en general hace más fácil e intuitivo su tratamiento, sino porque es aquella a la que estamos más acostumbrados. La información geográfica tiene una inherente naturaleza visual, ya que el espacio en sí es entendido de forma gráfica por el ser humano. Junto a esto, no debemos olvidar que la información geográfica se ha almacenado de forma tradicional de modo también visual, a través de mapas.

Al contrario que un mapa, que de por sí es de naturaleza gráfica, en un SIG trabajamos con datos de tipo puramente numérico, ya que es así como el ordenador puede manejarlos. Para poder presentar una utilidad similar a la de un mapa

en lo que a la presentación de la información respecta, un SIG debe incluir capacidades que generen representaciones visuales a partir de esos datos numéricos, aprovechando en la medida de lo posible las propias capacidades del medio informático en que se trabaja para hacer que estas representaciones sean más eficaces a la hora de transmitir información.

Es deseable igualmente que el SIG sea capaz de generar cartografía clásica, y que incorpore métodos para el diseño cartográfico y la creación de mapas impresos, pues estos no pierden su vigencia pese a la existencia de los SIG.

La visualización de la información geográfica se rige por los mismos conceptos y principios que se emplean para la confección de cartografía impresa, y estos deben ser conocidos por el usuario de SIG. A los conceptos tradicionales hay que sumar algunas ideas nuevas, ya que un SIG es capaz de generar representaciones más avanzadas (por ejemplo, representaciones tridimensionales o animaciones). A esto hay que sumar la presencia de un elemento característico y de gran importancia como es la elevada interactividad que toda representación gráfica lleva asociada dentro de un SIG, y que constituye una gran diferencia frente al carácter estático de la cartografía clásica.

Por todo ello, la visualización debe considerarse como un componente fundamental del sistema SIG en su concepción actual, y particularmente uno con especial interés desde el punto de vista del usuario directo de tecnologías SIG.

1.5.4. Tecnología

Incluimos en este elemento tanto el *hardware* sobre el que se ejecutan las aplicaciones SIG, como dichas aplicaciones, es decir el *software* SIG. Ambos forman un binomio tecnológico en el que encontramos diversas alternativas, y que se enriquece diariamente con la rápida evolución del mercado tecnológico.

En lo que a *hardware* respecta, es el elemento físico del sistema SIG, y conforma la plataforma sobre la que tiene lugar el trabajo con un SIG. La utilización de un SIG hoy en día se puede llevar a cabo en ordenadores personales o estaciones de trabajo, y ya sea de forma individual o en una arquitectura cliente–servidor más compleja. Estas últimas han cobrado importancia muy rápidamente en los últimos tiempos, especialmente en lo que al acceso a datos se refiere. Veremos más adelante como esto también ha tenido influencia en otros componentes del sistema SIG, principalmente en el factor organizativo.

Además de la propia plataforma, el *hardware* incluye una serie de periféricos para tareas más concretas. De uso habitual en el trabajo con SIG son los periféricos para entrada de datos geográficos y la creación de cartografía. Las tabletas digitalizadoras son la forma más habitual dentro del primer grupo (las veremos con más detalle en el apartado 6.5.1), mientras que *plotters* e impresoras son empleados para la creación cartográfica.

Más recientemente, la aparición de Sistemas de Navegación Global como el GPS (que pueden a su vez considerarse como otro tipo de periféricos) ha creado una parcela tecnológica con gran relación con los SIG, convirtiendo a estos en herramientas ideales para la gestión de los datos de dichos sistemas. Incluso, la combinación de SIG y GPS sobre un único elemento de *hardware* ha dado lugar a herramientas como los navegadores GPS, que han supuesto un hito no solo desde el punto de vista técnico, sino también desde un enfoque social, pues acercan las tecnologías SIG a usuarios no expertos.

Por su parte, el *software* es el encargado de operar y manipular los datos. El software SIG también ha sufrido una gran evolución, y bajo el paraguas de esa denominación encontramos

desde las aplicaciones clásicas que permiten visualizar, gestionar y analizar los datos geográficos, hasta herramientas más especializadas que se centran en alguno de estos campos, o bien componentes que pueden incluso pasar a formar parte de otras aplicaciones fuera del ámbito SIG, pero que puntualmente requieren algunas de sus funcionalidades, especialmente las relacionadas con la visualización.

1.5.5. Factor organizativo

El sistema SIG requiere una organización y una correcta coordinación entre sus distintos elementos. El factor organizativo ha ido progresivamente ganando importancia dentro del entorno SIG, a medida que la evolución de estos ha ido produciendo un sistema más complejo y un mayor número de intrarelaciones e interrelaciones entre los distintos componentes que lo forman.

Especialmente importante es la relación entre las personas que forman parte del sistema SIG, así como la relación de todos los elementos con los datos, sobre los cuales actúan de un modo u otro. Ello ha propiciado la aparición de, entre otros, elementos que pretenden estandarizar los datos y gestionar estos adecuadamente.

Cuando los SIG se encontraban en sus etapas de desarrollo iniciales y eran meras herramientas para visualizar datos y realizar análisis sobre ellos, cada usuario tenia sus propios datos con los cuales trabajaba de forma independiente del resto de usuarios, incluso si estos llevaban a cabo su trabajo sobre una misma área geográfica y estudiando las mismas variables. Hoy en día, la información no se concibe como un elemento privado de cada usuario, sino como un activo que ha de gestionarse, y del que deriva toda una disciplina completa. La aplicación de esta disciplina es la base de algunos de los avances más importantes en la actualidad, teniendo implicaciones no solo técnicas, sino también sociales en el ámbito de los SIG.

Asimismo, las necesidad de gestión de los datos y la propia complejidad de un SIG provocan ambas que no exista un perfil único de persona involucrada en el sistema SIG, sino varias en función de la actividad que desarrollen. Al usuario clásico de SIG se unen las personas responsables de gestionar las bases de datos, encargadas de diseñar la arquitectura de un SIG cuando este se establece para un uso conjunto por parte de toda una organización o grupo de mayor entidad. Dentro de las personas que participan en un SIG, el usuario directo es el eslabón último de una cadena que incluye igualmente a otros profesionales con roles bien distintos.

Incluso atendiendo únicamente a los usuarios, también entre estos existen diferentes perfiles, y las comunidades de usuarios no expertos juegan en la actualidad un importante papel en el mundo del SIG. Esta situación, a su vez, requiere elementos organizativos importantes. Con la popularización y bajo coste de las unidades GPS y la aparición de la denominada Web 2.0, el SIG ha llegado a usuarios no especializados, los cuales utilizan estas herramientas para la creación y uso de su propia cartografía, dentro de lo que se conoce como VGI (*Volunteered Geographic Information*[5]) [8]. El término *Neogeografía*, de reciente creación, hace referencia a este uso de los SIG y otras herramientas asociadas por parte de grupos de usuarios no especializados.

En definitiva, resulta necesario gestionar correctamente la complejidad del sistema SIG, y esta gestión se ha convertido ya en un elemento fundamental dentro del entorno SIG actual.

[5]Información geográfica creada voluntariamente

1.6. Resumen

En este capítulo hemos presentado los SIG como herramienta para el manejo general de información georreferenciada. Un SIG es un sistema compuesto por cinco piezas fundamentales: datos, tecnología, análisis, visualización y factor organizativo. Cada una de ellas cumple un papel determinado dentro del sistema SIG, el cual se caracteriza fundamentalmente por su naturaleza integradora.

Existen otras herramientas y tecnologías que pueden en principio asemejarse a los SIG, pero que realmente no comparten con estos su capacidad de integrar bajo un marco común una serie completa de elementos y disciplinas, siendo esta la verdadera propiedad que define a los SIG.

Todo el conjunto de conocimientos sobre los cuales se asientan los SIG conforman la denominada Ciencia de la Información Geográfica. Bajo esta denominación se recogen todos los temas a tratar en esta obra.

Capítulo 2

Historia de los SIG

2.1. Introducción

El desarrollo sufrido por los SIG desde sus orígenes hasta nuestros días es enorme. La popularización de las tecnologías y los esfuerzos de desarrollo llevados a cabo por un amplio abanico de ciencias beneficiarias de los SIG, todos han contribuido a redefinir la disciplina e incorporar elementos impensables entonces. No obstante, los componentes principales que identifican el núcleo principal de un SIG se mantienen a lo largo de todo ese desarrollo, y es su aparición la que define el momento inicial en el que podemos situar el origen de los SIG.

Este momento tiene lugar al inicio de la década de los sesenta, como resultado de dos factores: la necesidad creciente de información geográfica y de una gestión y uso óptimo de esta, y la aparición de los primeros computadores. Es en ese contexto donde nacen los SIG, con objeto de dar respuesta a esa necesidad utilizando para ello las nuevas herramientas disponibles.

2.2. Los orígenes

Las bases para la futura aparición de los SIG las encontramos algunos años antes de esa década de los sesenta, con el desarrollo de nuevos enfoques en cartografía que parecen predecir las necesidades futuras que un manejo informatizado de esta traerá. Los trabajos desarrollados por John K.Wright en la Sociedad Geográfica Americana, en especial la publicación de su obra *Elements of Cartography* en 1953, son particularmente importantes. Obras como esta van ampliando el campo de la geografía cuantitativa hasta que este alcanza un nivel donde puede plantearse, una vez que la informática alcanza una cierta madurez, la unión de ambas disciplinas.

La primera experiencia relevante en esta dirección la encontramos en 1959, cuando Waldo Tobler define los principios de un sistema denominado MIMO (map in–map out) con la finalidad de aplicar los ordenadores al campo de la cartografía. En él, establece los principios básicos para la creación de datos geográficos, su codificación, análisis y representación dentro de un sistema informatizado. Estos son los elementos principales del *software* que integra un SIG, y que habrán de aparecer en todas las aplicaciones desarrolladas desde ese momento.

El primer Sistema de Información Geográfica formalmente desarrollado aparece en Canadá, al auspicio del Departamento Federal de Energía y Recursos. Este sistema, denominado CGIS (Canadian Geographical Information Systems), fue desarrollado a principios de los 60 por Roger Tomlinson, quien dio forma a una herramienta que tenía por objeto el manejo de los datos del inventario geográfico canadiense y su análisis para la gestión del territorio rural. El desarrollo de Tomlinson es pionero en este campo, y se con-

sidera oficialmente como el nacimiento del SIG. Es en este momento cuando se acuña el término, y Tomlinson es conocido popularmente desde entonces como «el padre del SIG».

La aparición de estos programas no solo implica la creación de una herramienta nueva, sino también el desarrollo de técnicas que hasta entonces no habían sido necesarias. La más importante de ellas es la codificación y almacenamiento de la información geográfica, un problema en absoluto trivial que entonces era clave para lograr una usabilidad adecuada del *software*. El trabajo de Guy Morton con el desarrollo de su *Matriz de Morton*[1] juega un papel primordial[9], superando las deficiencias de los equipos de entonces, tales como la carencia de unidades de almacenamiento con capacidad de acceso aleatorio, que dificultaban notablemente el manejo y análisis de las bases de datos.

Simultáneamente a los trabajos canadienses, se producen desarrollos en Estados Unidos, en el seno del Harvard Laboratory, y en el Reino Unido dentro de la Experimental Cartography Unit. Ambos centros se erigen también como principales desarrolladores de *software* para la producción, manejo y análisis de información geográfica durante aquellos años.

En el Harvard Laboratory, ve la luz en 1964 SYMAP, un aplicación que permitía la entrada de información en forma de puntos, líneas y áreas, lo cual se corresponde a grandes rasgos con el enfoque que conocemos hoy en día como *vectorial*. En la imagen 2.1 puede verse que los resultados cartográficos de este *software* son aún de poca calidad. No obstante, el interés que despertaron las novedosas capacidades del programa para la generación de cartografía impulsó el desarrollo posterior y la evolución hacia sistemas más avanzados.

En 1969, utilizando elementos de una versión anterior de SYMAP, David Sinton, también en el Harvard Laboratory, desarrolla GRID, un programa en el que la información es almacenada en forma de cuadrículas. Hasta ese momento, la estructura de cuadrículas regulares era solo utilizada para las salidas de los programas, pero no para la entrada y almacenamiento de datos. Son los inicios de los Sistemas de Información Geográfica *ráster*[2].

SYMAP evoluciona y nuevos programas aparecen, tales como SYMVU (Figura 2.2), con capacidad de representación tridimensional, o CALFORM, con nuevas capacidades de representación y de generación de resultados impresos. GRID da lugar a IMGRID (Interactive Manipulation GRID), que sentará la base para el trabajo de Dana Tomlin con su paquete MAP, el cual incluye todos los elementos que hoy en día son imprescindibles para el análisis ráster (y que veremos en el capítulo 15)

Si la década de los sesenta es la de los pioneros y las primeras implementaciones, la de los setenta es la de la investigación y el desarrollo. A partir de los SIG primitivos se va dando forma a un área de conocimiento sin duda con gran futuro, y se elabora una base sólida de conocimiento y de herramientas aptas para un uso más genérico. Sin haber entrado aún en la época del uso masivo y generalizado, los primeros paquetes comienzan a distribuirse y pasan a incorporarse a la comunidad cartográfica, lejos ya de ser el producto de unos pocos pioneros.

A partir de este punto, el campo de los SIG recorre sucesivas etapas hasta nuestros días (Figura 2.3), evolucionando muy rápidamente ante la in-

[1] Veremos con algo más de detalle este concepto en el capítulo 5

[2] los conceptos de SIG ráster y vectorial se tratan extensamente en el capítulo 5. No te preocupes si ahora no comprendes completamente qué representa cada uno de ellos y qué los diferencia.

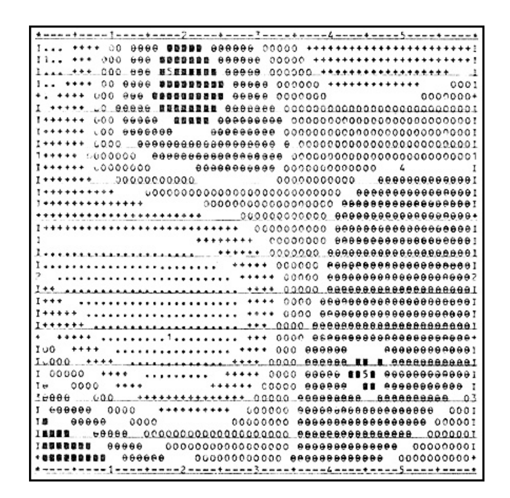


Figura 2.1: Aspecto de un mapa generado con SYMAP

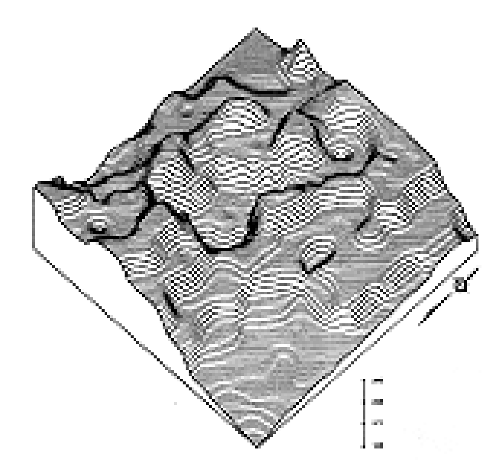


Figura 2.2: Representación tridimensional creada con SYMVU

fluencia de numerosos factores externos. Desde este punto, vamos a estudiar cómo esos factores han ido a su vez evolucionando y cómo su influencia ha condicionado el rumbo seguido por los SIG. Distinguiremos los siguientes elementos:

- La evolución del SIG como **disciplina**. Cómo ha cambiado la presencia social de los SIG y su relación con otras disciplinas científicas, tanto influenciándolas como siendo influenciado por ellas.
- La evolución de la **tecnología**. Cómo ha variado el *software* SIG, así como los ordenadores, periféricos y elementos informáticos de los que depende para su funcionamiento.
- La evolución de los **datos**. Cómo ha cambiado la generación de datos, su almacenamiento, y cómo esto ha condicionado el desarrollo de nuevas soluciones para su manejo.
- La evolución de las **técnicas y formulaciones**. Desde los elementos básicos de la cartografía cuantitativa, cómo se han desarrollado nuevos conceptos, enfoques, teorías o ramas de conocimiento, que han dejado su huella en la evolución de los SIG.

2.3. La evolución de los SIG como disciplina

Como hemos visto, los SIG eran en origen una mera combinación de elementos de cartografía cuantitativa, enlazados con los sistemas informáticos de la época. Se trataba de un territorio propio de cartógrafos y geógrafos que intentaban adaptar sus conocimientos y necesidades a las tecnologías que por aquel entonces comenzaban a surgir. Desde aquellos orígenes, los cambios han sido muy importantes, y se han incorporado al ámbito de los SIG un gran número de otras disciplinas cuya aportación e influencia puede ser equivalente o incluso superior a la de la cartografía o la geografía.

La utilización del término «geográfico» para denominar a estos sistemas de manejo de información ha hecho que, tradicionalmente, y a falta de una parcela de conocimiento propia bien delimitada, haya recaído en la geografía la tarea docente e investigadora relacionada con los SIG. No obstante, y dada la multidisciplinaridad del ámbito y su uso por grupos muy distintos hoy en día, no es necesariamente este el mejor en-

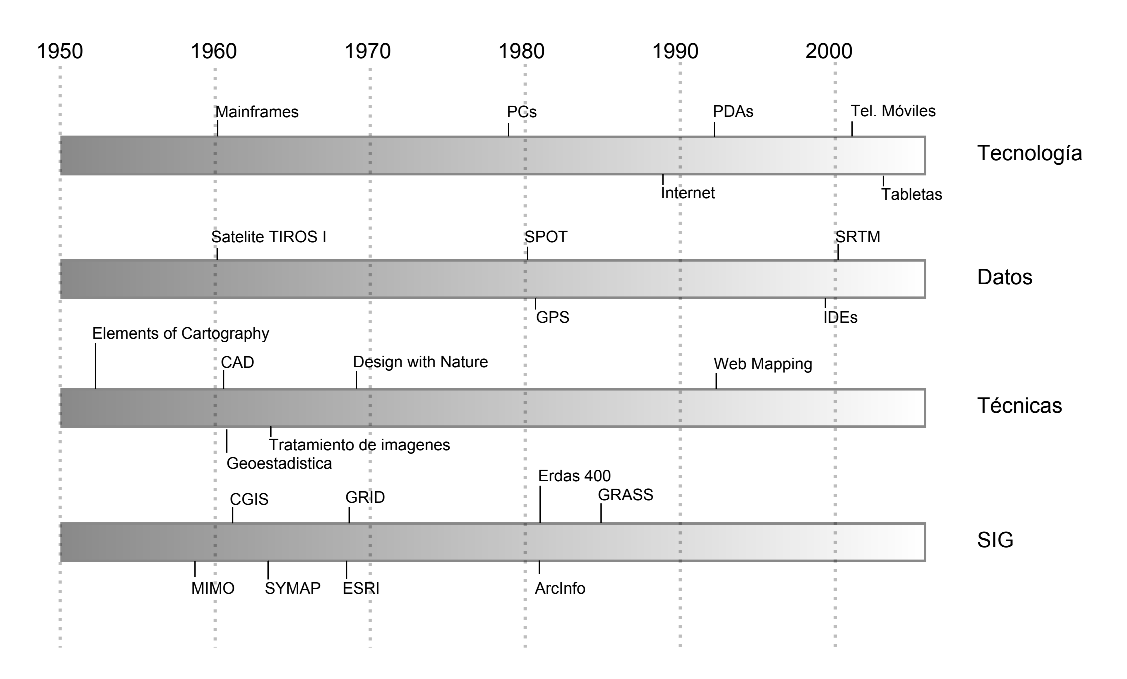


Figura 2.3: Esquema temporal de la evolución de los SIG.

foque [10]. En general, el conjunto de ciencias del medio y ciencias sociales han sabido todas ellas hacer uso de los SIG y aportar a estos los elementos propios de su ámbito.

Si bien los orígenes del SIG están íntimamente ligados a la gestión forestal o la planificación urbanística, son muchas otras las disciplinas que han jugado un papel importante. Un elemento sin duda clave es la sensibilización medioambiental, que obliga a un estudio del medio mucho más detallado. Coincidiendo con la etapa inicial del desarrollo de los SIG, empieza a aparecer una preocupación por el entorno que tiene consecuencias muy favorables para el desarrollo de todas las ciencias relacionadas, la gran mayoría de las cuales son o serán usuarias directas de SIG. El SIG comienza a integrarse paulatinamente en las tareas de gestión del medio, como un apoyo imprescindible a la hora de analizar este.

Al principio de la década de los setenta, siendo ya claro que los SIG son herramientas con gran futuro, aparecen no solo los esfuerzos de desarrollo y estabilización de la disciplina, sino todos los restantes que le dan entidad propia.

Así, a finales de septiembre de 1970, apenas media década después de que el CGIS fuera desarrollado, tiene lugar en Ottawa, Canada, el primer Simposio Internacional de Sistemas de Información Geográfica. La celebración de eventos similares será ya una actividad en constante aumento desde entonces.

Paralelamente, el SIG pasa a formar parte de los *curricula* universitarios y a constituirse en una disciplina bien diferenciada, al tiempo que el mercado editorial comienza a prestar atención a los SIG y aparecen obras clásicas que engloban toda la base conceptual de las herramientas modernas. Poco después, se crean las principales revistas especializadas, que recogen los avances y tendencias de una ciencia en muy rápido desarrollo.

En 1987 se empieza a publicar el *International Journal Of Geographical Information Systems*. Un año más tarde se funda en la Universidad Estatal de Nueva York, en Buffalo, la primera lista de distribución en Internet dedicada a los SIG, y arranca la publicación mensual *GIS World*.

Los productos del Harvard Laboratory se habían vendido a precios módicos a otros investigadores para financiar su propio desarrollo, pero sin gran afán comercial. La incorporación de los SIG al mercado y la aparición de una industria basada en ellos aparece poco después del inicio de estos, al final de los años sesenta. En 1969, Jack Dangermond, un integrante del propio Harvard Laboratory, funda junto a su esposa la empresa Environmental Systems Research Institute (ESRI), pionera y líder del sector hasta el día de hoy. La popularización de los SIG y su conversión en un elemento de consumo es debida también en gran medida a la labor de ESRI dentro del mercado y a su línea original de productos.

Esta popularización de la herramienta, acompañada de la disponibilidad creciente de ordenadores personales, hace que los SIG pasen de ser elementos al alcance de unos pocos a estar disponibles para todos los investigadores en una gran variedad de ámbitos. La multidisciplinaridad de los SIG como útiles de trabajo para todas las ciencias del medio se ve reforzada a partir de este momento con continuas aportaciones por parte de estas y la aceptación del SIG como un elemento más dentro de innumerables campos de trabajo.

Surgen nuevas empresas en el mercado, y en 1985 aparece el primer SIG libre, GRASS (Geographic Resources Analysis Support System), siendo aún en la actualidad un referente dentro de su área. También en la década de los ochenta, comienzan a perder sentido los primeros desarrollos con los que comenzó el SIG, y programas tales como CGIS no se encuentran ya en condiciones de competir en el mercado, que se desarrolla muy rápidamente y va creando soluciones adaptables.

En este sentido, es reseñable el hecho de que los SIG dejan de ser sistemas completos y pasan a ser plataformas adaptables sobre las que construir soluciones particulares. Los SIG se convierten en herramientas base para todo ese gran conjunto de disciplinas beneficiarias, cada una de las cuales adapta y particulariza estos a la medida de sus necesidades.

Con el paso del tiempo, los SIG van confluyendo y los diversos enfoques se unen para constituir una base útil sobre la que construir nuevos desarrollos. Los SIG ráster incluyen cada vez más elementos vectoriales, los SIG vectoriales cada vez más elementos ráster, y en ambos se van implementando formulaciones que trabajan con ambos formatos de almacenamiento y los combinan. De forma similar, los procesos para análisis de imágenes van ganando su espacio dentro de los SIG generales, aunque no dejan de existir aplicaciones específicas en este terreno.

Por último, respecto a su presencia social, en nuestros días los SIG han pasado de elementos restringidos para un uso profesional a ser elementos de consumo y estar presentes en nuestra vida diaria. Un ejemplo de ello es la aparición de servicios como *Google Maps* y la multitud de aplicaciones con interfaces Web basadas en él que permiten acceder a información geográfica de toda clase. De la mano también de *Google*, *Google Earth* es otra aplicación popular que no está restringida al uso profesional. Estas aplicaciones acercan los SIG a usuarios no especializados, dándoles la posibilidad de utilizarlos y aprovechar parte de sus capacidades.

La popularización de los navegadores GPS, que incorporan tanto elementos de representación como de análisis propios de los SIG, es otro buen ejemplo.

2.4. La evolución de la tecnología

La tecnología sobre la que se basan los SIG es clave para entender todo lo relacionado con ellos, especialmente su evolución a lo largo del tiempo. Desde los primeros SIG, muy lejos del alcance de un usuario medio, hasta las aplicaciones de escritorio o los elementos derivados de los SIG que son de uso habitual hoy en día, se ha producido un cambio enorme que, como cabe esperar, es paralelo al que la propia tecnología ha sufrido.

Tres son los bloques principales del desarrollo informático con una influencia más marcada en el campo de los Sistemas de Información Geográfica [11]:

- **Salidas gráficas**. Sin las capacidades de representación gráficas de hoy en día, puede parecernos imposible el uso de un SIG, ya que, aunque los procesos de análisis son una parte imprescindible y definitoria del mismo y pueden llevarse a cabo sin necesidad de visualización, esta visualización es una herramienta fundamental de un SIG. No obstante, tanto los primeros ordenadores como las primeras impresoras dedicadas a la impresión de mapas carecían de dichas capacidades. Como puede verse en la figura 2.1, las representaciones en esos tiempos se basaban en el uso de caracteres y no en gráficos puramente dichos.
  La evolución de las capacidades gráficas, intensa desde esos inicios hasta nuestros días y aún muy activa, ha sido seguida de cerca por los SIG, que progresivamente van incorporando mejoras tanto en la representación en pantalla como en la generación de mapas impresos.
- **Almacenamiento y acceso de datos.** Desde el inicio, el almacenamiento y acceso de datos ha sido un problema clave en el cual se han producido grandes avances. Por una parte, los problemas asociados a los grandes volúmenes de información. Por otra, los relacionados con la lectura de estos, que ha de poder realizarse de forma fluida a pesar a dicho volumen. A medida que han ido aumentando las capacidades de almacenamiento y lectura, ha ido aumentando paralelamente el tamaño de los datos manejados, así como los soportes utilizados para ellos, y esta evolución paralela ha de continuar y condicionar la forma que adopten los SIG.
- **Entrada de datos**. Los datos geográficos utilizados en los primeros años de los SIG eran datos en papel que se digitalizaban y almacenaban mecánicamente en tarjetas perforadas en un único proceso mecánico. Hoy en día, y aunque veremos que las fuentes de datos han sufrido por su parte una gran evolución, sigue siendo necesaria la digitalización de una gran cantidad de datos. Desde esos sistemas mecánicos de tarjetas hasta los modernos equipos, la aparición de *scanners* de gran precisión y técnicas de digitalización automáticas, entre otros, ha cambiado completamente el ámbito de la entrada de datos para su uso en un SIG.

Además del avance de estos factores, la evolución general de los ordenadores afecta a todos los elementos de *software* que se ejecutan sobre ellos. De las grandes computadoras se pasa a los ordenadores personales, y los programas tales como los SIG realizan también esa transición de una a otra plataforma.

La elaboración y análisis de cartografía se convierte a finales de los años ochenta en una tarea que puede ya llevarse a cabo en equipos personales (PC) de bajo coste, lejos de las grandes máquinas y equipos dedicados de alto coste.

En 1978, la recientemente creada empresa ERDAS adapta para el PC un *software* de análisis de imágenes denominado IMGGRID, y comienza a distribuir este junto con un *hardware* relativamente asequible para uso personal. El ERDAS

400 System se convierte así en el primero de su clase con esas características.

Paralelamente, ArcInfo, de la compañía ESRI, se convierte en 1981 en el primer SIG que alcanza el ámbito de los ordenadores personales. Será también un producto de esta compañía, ArcView, el que en 1991 pase a popularizar el SIG como herramienta de escritorio.

A mitad de los 80, ArcInfo y ERDAS comienzan a distribuirse de forma conjunta en un producto comercial que integra el análisis vectorial con el tratamiento de imágenes dentro del entorno de un PC.

La evolución de las plataformas no se detiene ahí. Las tendencias actuales apuntan a llevar los SIG de forma genérica a plataformas móviles tales como teléfonos o tabletas, especialmente indicadas para la toma de datos en campo. La combinación de estos últimos con las tecnologías de posicionamiento global como el GPS se demuestra altamente práctica en este aspecto.

Elementos de SIG se incluyen también en los navegadores GPS cada día más populares, confirmando la tendencia de adaptar los SIG a los dispositivos portátiles, tanto para el análisis como para la consulta de la información geográfica.

La aparición de Internet es un hecho que ha modificado todos los aspectos de la sociedad actual, estén relacionados o no con ámbito científico. Los SIG no son, como cabe esperar, una excepción a esto, e Internet ha jugado un papel decisivo en redefinir el concepto de SIG que hoy conocemos.

El nacimiento de la World Wide Web (WWW) puede establecerse a finales de 1989, pero no será hasta 1993 cuando empiece a utilizarse directamente para actividades relacionadas con los SIG o la distribución de cartografía. En esta fecha aparece *Xerox PARC*, el primer servidor de mapas. *Mapserver*, uno de los principales servidores de cartografía en la actualidad, aparece a mediados de 1997.

El primer atlas digital en línea es el Atlas Nacional de Canadá, que se encuentra disponible desde 1994. Otros como MultiMap o MapQuest, que alcanzan gran popularidad, aparecen en 1996 y establecen la línea a seguir por sucesivos servicios de Internet relacionados con la información geográfica.

En 2005 aparece Google Maps, que además de ofrecer servicios de cartografía permite desarrollar nuevas aplicaciones sobre dichos servicios a través de una interfaz de programación abierta y documentada. Los conceptos de la Web 2.0 se adaptan así al ámbito de los SIG. El número de ideas y funcionalidades basados en Google Maps crece exponencialmente desde prácticamente su nacimiento, extendiendo la tecnología SIG a campos casi insospechados y muy distintos de los que originalmente constituían el ámbito de uso de los SIG.

2.5. La evolución de los datos

Los datos son el elemento principal del trabajo dentro de un SIG. Sin ellos, no tiene sentido un Sistema de Información Geográfica. Esta relación entre los datos y los elementos de *software* y *hardware* empleados en su manejo ha ejercido una notable influencia en el desarrollo de las tecnologías SIG y, recíprocamente, estas han definido el marco de trabajo para los avances en los tipos de datos.

Los datos geográficos con los que se trabajaba en los primeros SIG provenían de la digitalización de cartografía impresa. Las primeras bases de datos geográficas contenían mapas escaneados y elementos digitalizados en base a estos.

A partir de este punto, van apareciendo nuevas fuentes de datos cuya estructura es más adecuada para su tratamiento informatizado, y al

tiempo que los SIG se adaptan a estas, surge una relación bidireccional que resulta beneficiosa para ambos.

Un avance primordial en este sentido lo constituye el lanzamiento de los primeros satélites de observación terrestre. Las técnicas existentes para la toma de fotografías aéreas, desarrolladas principalmente con fines militares durante la Primera Guerra Mundial, pasan a ser aplicadas a escala global con la aparición de satélites destinados a estos efectos.

El 1960, el primer satélite de observación meteorológico, el *TIROS I*, es lanzado al espacio. Dos años después, Rusia lanza su satélite *Kosmos*, y en 1974 el primer prototipo del satélite SMS–1 es puesto en órbita.

Otros hitos importantes son los lanzamientos de los satélites LANDSAT 2 y 7 en 1975 y 1999 respectivamente, cuyos productos son ambos de uso muy extendido (como veremos en el capítulo 6).

El 1980 se funda SPOT, la primera compañía mundial en ofrecer con carácter comercial imágenes procedentes de satélite para toda la superficie terrestre. A este hecho le seguiría el lanzamiento de un buen número de nuevos satélites con o sin fines comerciales. Los productos de la teledetección pasan a constituir una fuente de negocio, al tiempo que se incorporan como elementos básicos del análisis geográfico.

Las tecnologías de posicionamiento y localización son otra fuente de datos de primer orden. En 1981, el sistema GPS pasa a ser plenamente operativo, y en 2000 se amplía la precisión de este para uso civil. Este último hecho aumenta la penetración de la tecnología, pudiendo ya ser empleado el sistema para el desarrollo de elementos como navegadores GPS u otros productos derivados, hoy en día de uso común.

Al igual que las aplicaciones, los distintos tipos de datos geográficos digitales se van asentando y popularizando, recibiendo progresivamente más atención y medios. El Servicio Geográfico Estadounidense (USGS) publica en 1976 los primeros Modelos Digitales de Elevaciones (MDE), en respuesta a la gran importancia que este tipo de dato tiene dentro del nuevo contexto del análisis geográfico.

La evolución de los datos de elevación con cobertura global llega a un punto histórico en el año 2000 con la *Shuttle Radar Topographic Mission* (SRTM). La SRTM es un proyecto conjunto dirigido entre la NASA y la National Imagery and Mapping Agency (NIMA), cuyo objetivo es ofrecer información altitudinal de un 80 % de la superficie terrestre a una resolución de un segundo de arco (aproximadamente, 30 metros).

La aparición de nuevas técnicas tales como el LiDAR (ver 6.4.2) abre nuevos caminos en cuanto a la precisión que puede obtenerse en la caracterización del terreno, posibilitando nuevos usos y análisis antes no planteados.

La evolución de los datos no es solo una evolución técnica, sino también de carácter social y organizativo. En la denominada *era de la información*, el papel de los datos es tenido cada vez más en cuenta, y los esfuerzos para coordinar la enorme cantidad de datos espaciales y sus numerosas procedencias se hacen cada vez más relevantes. Se empieza a entender que resulta necesario formular estrategias adecuadas para la gestión de los datos espaciales. Estas estrategias pasan por la creación de las denominadas *Infraestructuras de Datos Espaciales* (IDE).

El ejemplo más destacado de estas es la IDE Nacional de los Estados Unidos (NSDI)[12], surgida a raíz de la Orden Ejecutiva 12096, que fue promulgada en 1994 y tuvo una vital importancia en este ámbito. En Europa, la directiva INSPIRE[13], con fecha 14 de marzo de 2007, pretende la creación de una infraestructura similar.

Muchos de estos desarrollos y actividades se adhieren a las especificaciones establecidas por el *Open GIS Consortium* (OGC), un consorcio internacional fundado en 1994 para homogeneizar el empleo y difusión de los datos geográficos.

2.6. La evolución de las técnicas y formulaciones

Los problemas iniciales de los pioneros del SIG eran el desarrollo de los primeros programas —esto es, la mera implementación— y los relativos al almacenamiento y codificación de datos, como ya vimos. Las formulaciones de estos inicios eran las de la cartografía cuantitativa del momento, aún no muy desarrollada. Una vez que se implementan los primeros SIG y se da respuesta a las necesidades de análisis y gestión de datos espaciales que motivaron su aparición, comienza el proceso de desarrollar nuevas técnicas y planteamientos que permiten ir más allá en dicho análisis.

La cartografía cuantitativa sufre desde entonces un avance muy notable, arrastrada por las necesidades de los SIG en su propia evolución, y muchas disciplinas científicas desarrollan nuevas formulaciones que comienzan a tener como base los Sistemas de Información Geográfica. Algunas de ellas resultan especialmente relevantes y pasan a formar parte del conjunto habitual de herramientas y elementos de un SIG genérico.

Como indica [14], la mayoría de los avances de cierta importancia dentro del mundo de los SIG han venido motivadas por la necesidad de una utilización concreta o por la tecnología en sí, y pocas veces por el desarrollo puro de una teoría. No obstante, e independientemente de las razones que lo motiven, los SIG han servido como contexto ideal para dar cuerpo a estas teorías, y su historia debe considerarse de forma pareja.

Antes de que aparecieran los primeros SIG, los trabajos de algunos pioneros establecen bases que más tarde serán de gran importancia para otros avances. Junto con el ya citado *Elements of Cartography* de John K.Wright, los trabajos de Ian McHarg anticipan una forma de operar con los datos geográficos que más adelante va a convertirse en una constante del trabajo con estos dentro de un SIG. En su libro *Design with Nature* (1969), McHarg define los elementos básicos de la superposición y combinación de mapas, que, como veremos más adelante, son los que se aplican tanto en el análisis como en la visualización de las distintas *capas* de datos geográficos en un SIG.

Aplicaciones de esta índole, en las cuales se combinan diversos mapas temáticos, ya se habían llevado a cabo con anterioridad. McHarg, sin embargo, es el encargado de generalizarlas como metodologías de estudio y análisis geográfico, y asienta de esta manera los fundamentos que luego pasaran a formar parte de los SIG. Su trabajo tiene, además, un fuerte componente medioambiental, elemento que, como ya se ha dicho, es una de las razones que impulsan al desarrollo de los SIG como herramientas para una mejor gestión del medio.

El análisis cartográfico, aun así, arranca desde mucho antes de McHarg. En 1854, John Snow (1813–1858) realizó la que puede considerarse como una de las primeras experiencias cartográficas analíticas, al utilizar mapas de puntos para efectuar sus deducciones y localizar en Inglaterra la fuente de un brote de cólera.

Junto con la componente analítica, otros elementos de la práctica cartográfica evolucionan de manera similar. En 1819, Pierre Charles Dupin crea el primer mapa de coropletas para mostrar la distribución del analfabetismo en Francia, dando un gran salto cualitativo en el diseño cartográfi-

co, particularmente en un tipo de mapas de muy habitual creación dentro de un SIG.

Una vez que los SIG han hecho su aparición, entre los elementos que más han impulsado su desarrollo cabe destacar el gran avance en el estudio del relieve, de notable importancia por ser un elemento base para muchos otros análisis en un amplio abanico de ciencias afines. La orografía clásica, con un enfoque tradicionalmente sustentado en la geología y el análisis geomorfológico, va dando lugar a una ciencia cada vez más cuantitativa centrada en el análisis morfométrico del relieve. Trabajos como los de [15] sientan las bases para este tipo de análisis, que necesitan de un SIG para ser aplicados de forma efectiva.

De igual modo sucede con la geoestadística, una rama de la estadística que aparece de la mano del francés Georges Matheron a principio de los años sesenta. Las formulaciones geoestadísticas, hoy parte característica de los SIG, se desarrollan en esa época desde el punto de vista teórico, aunque no son aplicables para un uso real si no es con el uso de ordenadores, y pierden gran parte de su valor práctico si no se realiza esta tarea con el concurso de Sistemas de Información Geográfica.

En general, el desarrollo de la estadística encaminado a la adaptación de teorías y metodologías al ámbito espacial ha tenido un fuerte crecimiento en las últimas décadas, un hecho muy ligado a la aparición y evolución de los SIG. Uno de los hitos de este proceso es el desarrollo de [16], que extiende los modelos autoregresivos, de importancia clave para el análisis de la variación de series temporales, a los datos espaciales [17].

Otro hecho importante es la aparición de los primeros programa de diseño asistido por ordenador (CAD) , que coincide con la de los SIG, allá por el final de los años sesenta. Originalmente pensados para el diseño industrial, pronto pasan a ser utilizados para el diseño arquitectónico y la delineación de elementos geográficos, y sus conceptos son incorporados paulatinamente a los SIG. Hoy en día, y cada vez con más frecuencia, los SIG incorporan capacidades similares a los sistemas CAD, que permiten tanto la digitalización de cartografía con las herramientas propias del CAD como la creación de nuevos elementos geográficos. Asimismo, los formatos habituales de las aplicaciones CAD son soportados por gran número de SIG, existiendo una cierta interoperabilidad, no obstante muy mejorable.

El avance en el desarrollo de las aplicaciones CAD, y en general de las representaciones gráficas por ordenador, impulsó igualmente la aparición y evolución posterior de una nueva disciplina: la geometría computacional. Esta denominación se emplea por primera vez en 1975 [18], siendo hoy el nombre de una rama de la ciencia consolidada y en constante avance. Los algoritmos que componen la geometría computacional son la base sobre la que se fundamenta el análisis vectorial dentro de un SIG.

2.7. Resumen

A principios de los años sesenta, el creciente interés por la información geográfica y el estudio del medio, así como el nacimiento de la era informática, propiciaron la aparición de los primeros SIG.

Desde ese punto hasta nuestros días, los SIG han ido definiéndose en base a la evolución de la informática, la aparición de nuevas fuentes de datos susceptibles de ser utilizadas en el análisis geográfico —muy especialmente las derivadas de satélites—, y del desarrollo de disciplinas relacionadas que han contribuido a impulsar el desarrollo propio de los SIG.

Siendo en su origen aplicaciones muy específicas, en nuestros días los SIG son aplicaciones genéricas formadas por diversos elementos, cuya tendencia actual es a la convergencia en productos más versátiles y amplios.

Capítulo 3

Fundamentos cartográficos y geodésicos

3.1. Introducción

La característica principal de la información georreferenciada es que tiene una localización en el espacio, particularmente en el espacio terrestre. Esta localización se ha de dar por medio de unas coordenadas que la definan de forma adecuada, lo cual implica la necesidad de establecer un sistema en base al cual expresar dichas coordenadas.

Si medimos un dato de temperatura necesitamos un sistema de medición conocido, sin el cual el dato de temperatura en sí carece de valor y significado. Así, no es lo mismo decir que una temperatura es de 25 grados Celsius o que es de 25 grados Fahrenheit. Del mismo modo, si a esa temperatura le queremos asociar algún tipo de información espacial (por ejemplo, el punto exacto en el que fue medida), debemos establecer un sistema que permita dar sentido a las mediciones que realicemos, y que posteriormente nos sirva para interpretar los valores de las coordenadas y poder saber con exactitud *dónde* está el punto al que estas hacen referencia.

El establecimiento de un sistema de referencia en el que expresar la situación de un punto dado no es en absoluto una tarea sencilla, y requiere el conocimiento de abundantes conceptos previos que van desde ideas físicas hasta complejos desarrollos matemáticos y geométricos. Los avances en este campo han sido constantes desde la antigüedad, y esta evolución es la que ha permitido que en la actualidad se puedan obtener resultados altamente precisos en el trabajo con información georreferenciada. Gran parte de lo que podemos hacer en un SIG carecería de sentido si no se dispusiera de metodologías bien desarrolladas para el establecimiento de sistemas de referencia.

La geodesia es la ciencia encargada de proveer el marco teórico en el que fundamentar todo lo anterior, y es una disciplina compleja con diversas ramas de estudio. Todas ellas responden al objetivo básico de estudiar la forma de la Tierra, ya que debemos saber cómo es la Tierra para poder localizar puntos sobre su superficie. La determinación de la forma y dimensiones de la Tierra es tarea de la denominada *geodesia esferoidal*, cuyo cometido coincide con el del concepto clásico de geodesia, esto es, la definición de la figura terrestre. No obstante, en la actualidad encontramos otras ramas como la *geodesia física*, encargada de analizar el campo gravitatorio terrestre y sus variaciones, o la *astronomía geodésica*, que utiliza métodos astronómicos para la determinación de ciertos elementos geodésicos muy importantes que veremos más adelante. En conjunto, todas estas ramas dan forma a una serie de métodos y conceptos que son los que van a permitir la utilización rigurosa de coordenadas.

La necesidad del estudio geodésico surge por el hecho de que la Tierra no es plana, y cuando el territorio que pretendemos estudiar es lo su-

ficientemente extenso, la curvatura de la Tierra no puede ignorarse. Este es el caso que vamos a encontrar cuando trabajemos con un SIG, y es por ello que los SIG implementan los elementos necesarios para poder efectuar un manejo de la información geográfica riguroso y acorde con los conceptos de la geodesia.

Vimos en el primer capítulo de esta parte que existen otras aplicaciones que trabajan con información georreferenciada, entre las cuales estaban los programas de diseño asistido por ordenador (CAD). Decíamos entonces que una de las principales limitaciones de estos era su mala disposición al trabajo con zonas extensas, ya que han sido diseñados para operar con zonas de unas dimensiones reducidas.

Cuando un arquitecto diseña el plano de una casa con una aplicación CAD, no necesita emplear los conceptos de la geodesia, puesto que a esa escala la forma de la Tierra no tiene relevancia. Aun prescindiendo de ella, puede expresar las coordenadas de los distintos elementos (un muro, un pilar, etc.) con la suficiente precisión y corrección como para que luego pueda construirse esa casa. Sin embargo, cuando un usuario de SIG estudia la cuenca vertiente de un río o la distribución de población en las regiones de un país, o bien analiza las rutas migratorias de un ave entre dos continentes, los conceptos de la geodesia resultan fundamentales.

En la actualidad, los SIG han hecho que la información geográfica tenga en muchos casos carácter global y cubra grandes extensiones o incluso la totalidad del planeta. Esto obliga más que nunca a hacer hincapié en los fundamentos geodésicos que resultan básicos para que toda esa información pueda manejarse correctamente, siendo de interés para cualquier usuario de SIG, con independencia de su escala de trabajo.

Otro aspecto básico a la hora de trabajar en un SIG son las denominadas *proyecciones cartográficas*. Estas permiten transformar las coordenadas sobre la superficie curva de la Tierra en coordenadas sobre una superficie plana. Esto es necesario para poder representarlas en un soporte plano tal como puede ser un mapa o la pantalla del ordenador, así como para poder analizarlas de forma más simple.

Con los elementos de la geodesia y las proyecciones cartográficas ya podemos elaborar cartografía y estamos en condiciones de trabajar con la información georreferenciada. No obstante, existen ciertos conceptos relativos a esa cartografía que resultan de suma importancia y deben conocerse antes de abordar esas tareas. El más importante de ellos es la escala, es decir, la relación entre el tamaño real de aquello que representamos y su tamaño en la representación, la cual constituye un factor básico de toda información cartográfica.

La escala condiciona a su vez la aparición de otra serie de ideas y de procesos asociados, como por ejemplo la generalización cartográfica. Esta engloba los procedimientos que permiten que a cada escala se represente la información de la forma más adecuada posible, maximizando el valor de dichas representaciones. Aunque tanto la escala como la generalización cartográfica son conceptos muy vinculados a las propias representaciones visuales de la información geográfica, y este libro contiene una parte dedicada específicamente a la visualización, se trata de conceptos cartográficos fundamentales y por ello se incluyen en este capítulo, ya que resultan necesarios incluso si se trabaja con datos georreferenciados sin visualización alguna de estos.

3.2. Conceptos geodésicos básicos

A la hora de definir la forma y dimensiones de la Tierra, la geodesia plantea modelos que

puedan recoger la complejidad natural de la superficie terrestre y expresarla de una forma más simple y fácil de manejar.

Con estos modelos, uno de los objetivos principales de la geodesia es establecer un sistema de referencia y definir un conjunto de puntos (conocidos como *vértices geodésicos*) cuyas coordenadas en dicho sistema sean conocidas con una precisión elevada. Posteriormente, y en base a esos puntos, los cuales forman una *red geodésica*, se pueden calcular las coordenadas de cualquier punto en el sistema de referencia definido.

Los vértices geodésicos se establecen por triangulación a partir de un punto único determinado por métodos astronómicos. En función de la longitud de los lados de los triángulos empleados en dicha triangulación, tenemos redes de mayor o menor precisión.

Veamos ahora cómo establecer los elementos necesarios para establecer ese sistema de referencia base y definir esos modelos de partida citados. A la hora de buscar un modelo al que asimilar la forma de la Tierra, existen dos conceptos básicos: el *elipsoide de referencia* y el *geoide*.

3.2.1. Elipsoide de referencia y geoide

El intento más básico de establecer un modelo de la forma de la Tierra es asimilar esta a una figura geométrica simple, la cual pueda expresarse mediante una ecuación matemática. Además de ser más sencilla de manejar, disponer de esta ecuación matemática permite la aplicación de conceptos geométricos, estableciendo así una base práctica para el trabajo con coordenadas y la definición de sistemas de referencia.

Desde la antigüedad, se han formulado numerosas hipótesis sobre la forma que la Tierra tenía, las cuales van desde suponer la Tierra plana a admitir la evidencia de que esta ha de tener forma esférica (o similar) si se atiende a diversos hechos como, por ejemplo, el movimiento circular de las estrellas o la existencia de horizonte.

En realidad, la Tierra no es una esfera perfecta, ya que su propia rotación ha modificado esa forma y ha provocado un achatamiento en los polos. Esta hipótesis fue ya planteada por Newton, y corroborada posteriormente con numerosas experiencias. No obstante, podemos seguir tratando de asimilar la forma de la Tierra a la de una superficie teórica, aunque no ya la de una esfera, sino la de lo que se denomina un *elipsoide*. Sobre un elipsoide, el radio de la Tierra ya no es constante, sino que depende del emplazamiento.

Suponer que la Tierra es una esfera no es una aproximación tan mala como puede parecer (las representaciones gráficas a las que estamos acostumbrados exageran habitualmente mucho el achatamiento del planeta), aunque el elipsoide es más preciso y necesario a la hora de elaborar cartografía de zonas no muy extensas. A gran escala, sin embargo, y para determinadas tareas, es habitual suponer la Tierra con una forma perfectamente esférica.

Como se muestra en la figura 3.1, un elipsoide viene definido por dos parámetros: el semieje mayor y el semieje menor.

En el caso de la Tierra estos se corresponderían con el radio ecuatorial y el radio polar respectivamente. La relación existente entre estas dos medidas define el grado de achatamiento del elipsoide. En particular, se establece un factor de achatamiento según

$$f = \frac{r_1 - r_2}{r_1} \qquad (3.2.1)$$

siendo r_1 el semieje mayor y r_2 el semieje menor.

El elipsoide es la forma geométrica que mejor se adapta a la forma real de la Tierra, y por tanto la que mejor permite idealizar esta, logrando un mayor ajuste.

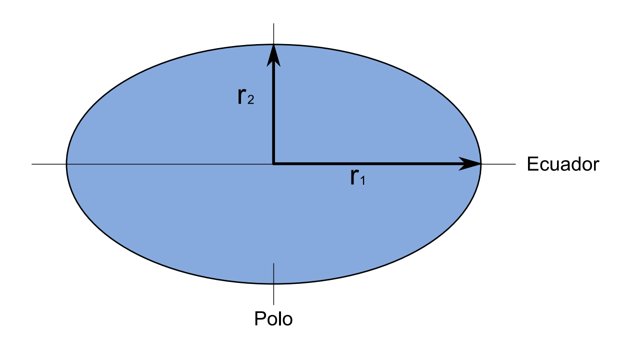


Figura 3.1: Parámetros que definen el elipsoide

Una vez que se dispone de una expresión teórica para la forma de la Tierra, el siguiente paso es la determinación de los parámetros que definen esta. En el caso de utilizar la esfera, hay que calcular su radio. En el caso de asumir el elipsoide como forma de referencia, deben determinarse las medidas de los semiejes menor y mayor.

Debido a la evolución histórica de la idea de elipsoide de referencia, las medidas de los semiejes que lo definen no son únicas. Es decir, no en todos lugares y en todas las circunstancias se emplea un mismo elipsoide caracterizado por unos valores r_1 y r_2 idénticos. Esto es debido principalmente al hecho de que un determinado elipsoide no se adapta de modo igualmente preciso a todas las regiones terrestres, y el elipsoide que proporciona un mejor ajuste para un área dada (por ejemplo, un continente o país) puede no ser el mejor en otra zona de la Tierra alejada de la primera.

A esto debe sumarse que los esfuerzos iniciales por determinar la forma de la Tierra y los parámetros del elipsoide de referencia fueron realizados en tiempos en los que la comunicación entre distintos puntos de la superficie terrestre no era la misma que hoy en día. Por ejemplo, los geodestas europeos de entonces realizaban un trabajo similar a sus colegas americanos, pero los datos con los que contaban eran bien distintos, pues las mediciones de cada grupo eran relativas a sus zonas de trabajo, al no resultar sencillo desplazarse a otras partes del planeta a realizar una labor similar.

De este modo, los geodestas de Europa tomaban sus datos y ajustaban a estos sus elipsoides, mientras que los de América hacían un trabajo similar y obtenían sus propios elipsoides. A la hora de establecer un elipsoide de referencia oficial, en cada zona (ya sea administrativa o geográfica) se tomaba el más idóneo, que no era el mismo en todas ellas.

Si añadimos las diferencias tecnológicas y metodológicas que también existían en el proceso de recogida y procesado de datos, es fácil comprender que tengamos una larga serie de elipsoides, cada uno de los cuales ha sido empleado de forma regular en un país o grupo de países, o incluso a escala continental, pero no a nivel global.

La tabla 3.1 muestra algunos de los elipsoides de uso más extendido en diversas partes del mundo, con sus correspondientes parámetros.

La necesidad de trabajar con un elipsoide global para todo el planeta es más reciente, pero desde hace ya casi un siglo resulta evidente que debe realizarse un esfuerzo por homogeneizar el uso de elipsoides, de tal modo que pueda trabajarse con una referencia internacional que facilite el uso de cartografía en las distintas zonas del pla-

Elipsoide	Semieje mayor	Semieje menor	$\frac{1}{f}$
Australian National	6378160.000	6356774.719	298.250
Bessel 1841	6377397.155	6356078.963	299.152
Clarke 1866	6378206.400	6356583.800	294.978
Clarke 1880	6378249.145	6356514.870	293.465
Everest 1956	6377301.243	6356100.228	300.801
Fischer 1968	6378150.000	6356768.337	298.300
GRS 1980	6378137.000	6356752.314	298.257
International 1924	6378388.000	6356911.946	297.000
SGS 85	6378136.000	6356751.302	298.257
South American 1969	6378160.000	6356774.719	298.250
WGS 72	6378135.000	6356750.520	298.260
WGS 84	6378137.000	6356752.314	298.257

Cuadro 3.1: Algunos elipsoides y sus parámetros característicos

neta. Como consecuencia de esto, surgen los primeros elipsoides *generales* (en contraste con los elipsoides *locales*), los cuales, además de buscar un ajuste óptimo, han de cumplir las siguientes características:

- El centro de gravedad terrestre y el del elipsoide deben coincidir.
- El plano ecuatorial terrestre y el del elipsoide deben coincidir.

El elipsoide WGS–84 es uno de los más empleados en la actualidad, ya que es el utilizado por el sistema GPS (apartado 6.6).

El geoide es la otra superficie de referencia, definida como la superficie tridimensional en cuyos puntos la atracción gravitatoria es constante. Se trata de una superficie equipotencial que resulta de suponer los océanos en reposo y a un nivel medio (el nivel es en realidad variable como consecuencia de las mareas, corrientes y otros fenómenos) y prolongar estos por debajo de la superficie terrestre. La particularidad del geoide reside en que en todos sus puntos la dirección de la gravedad es perpendicular a su superficie.

El geoide no es, sin embargo, una superficie regular como el elipsoide, y presenta protuberancias y depresiones, como puede observarse en la figura 3.2. La densidad de la Tierra no es constante en todos sus puntos, y ello da lugar a que el geoide sea una superficie irregular como consecuencia de las anomalías gravimétricas que dichas variaciones de densidad ocasionan.

Lógicamente, el elipsoide, por su naturaleza más simple, no puede recoger toda la variabilidad del geoide, por lo que estas dos superficies presentan diferencias, cuyo máximo es generalmente del orden de ± 100 metros. Estas diferencias se conocen como *alturas geoidales*.

Al igual que en el caso de los elipsoides, existen diversos geoides de referencia, y estos no son constantes en el tiempo, sino que evolucionan para adaptarse a las modificaciones que tienen lugar sobre la superficie terrestre.

La figura 3.3 muestra una comparación esquemática entre las tres superficies: superficie real de la Tierra, geoide y elipsoide.

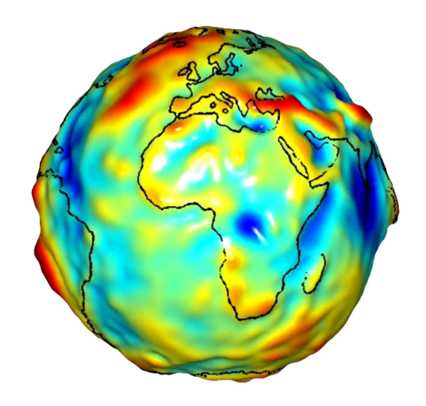


Figura 3.2: Representación gráfica del geoide (Fuente: Misión GRACE (NASA)).

Figura 3.3: Tres superficies fundamentales: superficie real de la Tierra, geoide y elipsoide (Adaptado de Wikipedia).

3.2.2. El datum geodésico

Cuando se trabaja con un elipsoide general, este, como se ha dicho, se sitúa de tal modo que tanto la posición de su centro de gravedad como su plano ecuatorial coincidan con los terrestres. Por el contrario, cuando el elipsoide es local, estas propiedades no han de cumplirse necesariamente, y el elipsoide a solas resulta insuficiente ya que carecemos de información sobre su posicionamiento con respecto a la superficie terrestre.

Surge así el concepto de *datum*, que es el conjunto formado por una superficie de referencia (el elipsoide) y un punto en el que «enlazar» este al geoide. Este punto se denomina *punto astronómico fundamental* (para su cálculo se emplean métodos astronómicos), o simplemente *punto fundamental*, y en él el elipsoide es tangente al geoide. La altura geoidal en este punto es, como cabe esperar, igual a cero. La vertical al geoide y al elipsoide son idénticas en el punto fundamental.

Para un mismo elipsoide pueden utilizarse distintos puntos fundamentales, que darán lugar a distintos datum y a distintas coordenadas para un mismo punto.

3.3. Sistemas de coordenadas

Disponiendo de un modelo preciso para definir la forma de la Tierra, podemos establecer ya un sistema de codificar cada una de las posiciones sobre su superficie y asignar a estas las correspondientes coordenadas. Puesto que la superficie de referencia que consideramos es un elipsoide, lo más lógico es recurrir a los elementos de la geometría esférica y utilizar estos para definir el sistema de referencia. De ellos derivan los conceptos de latitud y longitud, empleados para establecer las *coordenadas geográficas* de un punto.

No obstante, la geometría plana resulta mucho más intuitiva y práctica que la geometría esférica para realizar ciertas tareas, y a raíz de esto surgen las *proyecciones cartográficas*, que tratan de situar los elementos de la superficie del elipsoide sobre una superficie plana, y que son los que se emplean para la creación de cartografía. Al aplicar una proyección cartográfica, las coordenadas resultantes son ya coordenadas cartesianas.

Ambas formas de expresar la posición de un punto se utilizan en la actualidad, y las veremos con detalle en esta sección.

3.3.1. Coordenadas geográficas

El sistema de coordenadas geográficas es un sistema de coordenadas esféricas mediante el cual un punto se localiza con dos valores angulares:

- la **latitud** ϕ es el ángulo entre la línea que une el centro de la esfera con un punto de su superficie y el plano ecuatorial. Las líneas formadas por puntos de la misma latitud se denominan *paralelos* y forman círculos concéntricos paralelos al ecuador. Por definición, la latitud es de 0°en el ecuador, que divide el globo en los hemisferios norte y sur. La latitud puede expresarse especificando si el punto se sitúa al norte o al sur, por ejemplo 24°, 21' 11" N, o bien utilizando un signo, en cuyo caso los puntos al Sur del ecuador tienen signo negativo.
- la **longitud** λ es el angulo formado entre dos de los planos que contienen a la línea de los Polos. El primero es un plano arbitrario que se toma como referencia y el segundo es el que, ademas de contener a la línea de los polos, contiene al punto en cuestión. Las líneas formadas por puntos de igual longitud se denominan *meridianos* y convergen en los polos.
  Como meridiano de referencia internacional se toma aquel que pasa por el observatorio de Greenwich, en el Reino Unido. Este divide a su vez el globo en dos hemisferios: el este y el oeste. La longitud puede expresarse especificando si el punto se sitúa al Este o al Oeste, por ejemplo 32°, 12' 43" E, o bien utilizando un signo, en cuyo caso los puntos al Oeste del meridiano de referencia tienen signo negativo.

En la figura 3.4 puede verse un esquema de los conceptos anteriores.

Las coordenadas geográficas resultan de gran utilidad, especialmente cuando se trabaja con grandes regiones. No obstante, no se trata de un sistema cartesiano, y tareas como la medición de áreas o distancias es mucho más complicada. Si bien la distancia entre dos paralelos es prácticamente constante (es decir, un grado de latitud equivale más o menos a una misma distancia en todos los puntos), la distancia entre dos meridianos no lo es, y varía entre unos 11,3 kilómetros en el Ecuador hasta los cero kilómetros en los polos, donde los meridianos convergen.

3.3.2. Proyecciones cartográficas

A pesar de su innegable utilidad y la potencia que nos brindan para la localización de cualquier punto sobre la superficie terrestre, un sistema de coordenadas esféricas tiene inconvenientes que no pueden obviarse. Por una parte, estamos más acostumbrados a la utilización de sistemas cartesianos en los cuales la posición de un punto se define mediante un par de medidas de distancia x e y. Por otro lado, si necesitamos crear una representación visual de la información cartográfica, lo habitual es hacerlo en una superficie plana, ya sea a la manera clásica en un pliego de papel o, usando las tecnologías actuales, en un dispositivo tal como una pantalla.

Por todo ello, se deduce que existe una necesidad de poder trasladar la información geográfica (incluyendo, por supuesto, la referente a su localización) a un plano, con objeto de poder crear cartografía y simplificar gran número de operaciones posteriores. El proceso de asignar una coordenada plana a cada punto de la superficie de la Tierra (que no es plana) se conoce como *proyección cartográfica.*

Más exactamente, una *proyección cartográfica* es la correspondencia matemática biunívoca entre los puntos de una esfera o elipsoide y sus

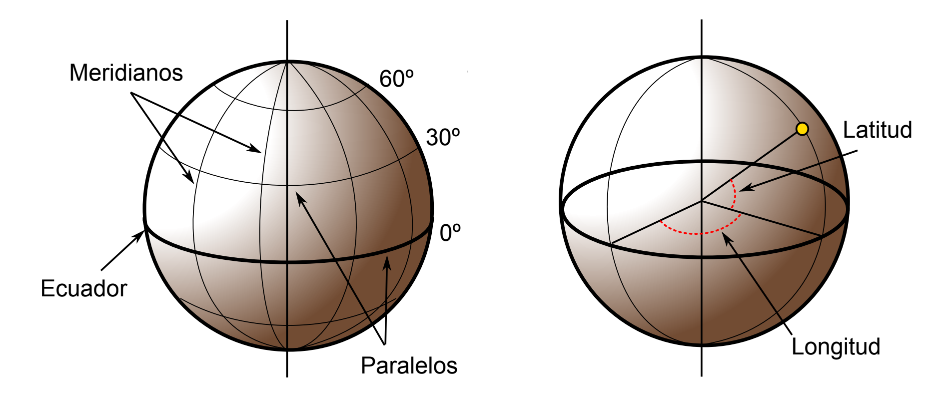


Figura 3.4: Esquema de los elementos del sistema de coordenadas geográficas.

transformados en un plano [19]. Es decir, una aplicación f que a cada par de coordenadas geográficas (ϕ, λ) le hace corresponder un par de coordenadas cartesianas (x, y), según

$$x = f(\phi, \lambda)\ ;\ y = f(\phi, \lambda) \tag{3.3.1}$$

De igual modo, las coordenadas geográficas puede obtenerse a partir de las cartesianas según

$$\phi = g(x, y)\ ;\ \lambda = g(x, y) \tag{3.3.2}$$

Se puede pensar que podemos obtener una representación plana de la superficie de una esfera o un elipsoide si tomamos esta y la extendemos hasta dejarla plana. Esto, sin embargo, no resulta posible, ya que dicha superficie no puede *desarrollarse* y quedar plana. Por ello, hay que buscar una forma distinta de relacionar los puntos en la superficie tridimensional con nuevos puntos en un plano.

La figura 3.5 muestra un esquema del concepto de proyección, esbozando la idea de cómo puede establecerse la correspondencia entre puntos de la esfera y del plano.

En ella vemos cómo el concepto de proyección se asemeja a la generación de sombras, ya que a partir de un foco se trazan las trayectorias de una serie de rayos que unen dicho foco con los puntos a proyectar, y después se determina el punto de contacto de esos rayos con la superficie plana. Aunque no todas las proyecciones siguen necesariamente este esquema, una parte de ellas sí que se fundamentan en un razonamiento similar a este, y el esquema mostrado sirve bien para entender el concepto y el paso de coordenadas de una superficie tridimensional a una bidimensional.

Veremos en los siguientes puntos las diferentes modificaciones que pueden introducirse sobre la forma anterior de proyectar, y que dan lugar a tipos distintos de proyecciones.

Puede apreciarse igualmente en la figura que se producen distorsiones al realizar la proyección. Es decir, que ciertas propiedades no se reproducen con fidelidad al pasar puntos desde la superficie curva al plano. Por ejemplo, la distancia entre los puntos A y B no es igual a la existente entre los puntos a y b. Con independencia de las características propias de la proyección, siempre existen distorsiones. Esto es así debido a que la esfera, como se ha dicho, no es desarrollable, mientras que el plano sí lo es, y por ello en el paso de coordenadas de uno a otra han de aparecen inevitablemente alteraciones.

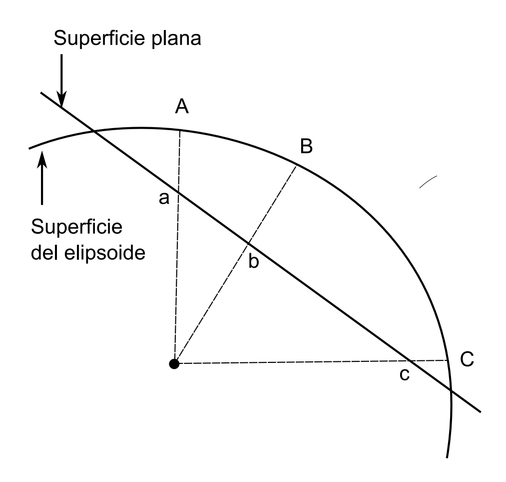


Figura 3.5: Esquema del concepto de proyección. A los puntos A, B y C sobre la superficie del elipsoide les asocian equivalentes a, b y c sobre un plano.

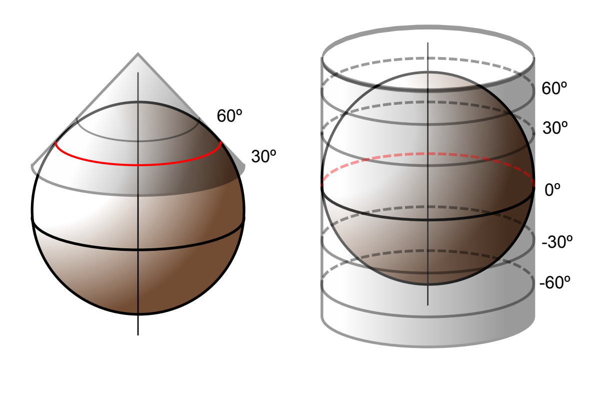


Figura 3.6: Esquema de una proyección cónica (izda.) y cilíndrica(dcha.)

Tipos de proyecciones

Las proyecciones se clasifican según la superficie sobre la que se proyectan los puntos. En el esquema de la figura 3.5, el plano de proyección es ya de por sí bidimensional. No obstante, puede realizarse la proyección sobre una superficie tridimensional, siempre que esta, a diferencia de la esfera, sí sea desarrollable. Es decir, que pueda «desenrollarse» y convertirse en un plano sin necesidad de doblarse o cortarse. Estas otras superficies pueden emplearse también para definir una proyección, de la misma forma que se hace con un plano.

Las superficies más habituales son el cono y el cilindro (junto con, por supuesto, el plano), las cuales, situadas en una posición dada en relación con el objeto a proyectar (esto es, la Tierra), definen un tipo dado de proyección. Distinguimos así los siguiente tipos de proyecciones:

- **Cónicas**. La superficie desarrollable es un cono (Figura 3.6), que se sitúa generalmente tangente o secante en dos paralelos a la superficie del elipsoide. En este último caso, la distorsión se minimiza en las áreas entre dichos paralelos, haciéndola útil para representar franjas que no abarquen una gran distancia en latitud, pero poco adecuada para representación de grandes áreas. Algunas de las proyecciones más conocidas de este grupo son la proyección cónica equiárea de Albers y la proyección conforme cónica de Lambert.
- **Cilíndricas**. La superficie desarrollable es un cilindro (Figura 3.6). Al proyectar, los meridianos se convierten en líneas paralelas, así como los paralelos, aunque la distancia entre estos últimos no es constante.
  En su concepción más simple, el cilindro se sitúa de forma tangente al ecuador (proyección normal o simple), aunque puede situarse secante y hacerlo a los meridianos (proyección transversa) o a otros puntos (proyección oblicua).
  La proyección de Mercator, la transversa de Mercator, la cilíndrica de Miller o la cilíndrica equiárea de Lambert son ejemplos relativamente comunes de este tipo de proyecciones.
- **Planas o azimutales**. La superficie desarrollable es directamente un plano. Según el esque-

ma de la figura 3.5, tenemos distintos tipos en función de la posición del punto de fuga.

- Gnómica o central. El punto de fuga se sitúa en el centro del elipsoide.
- Estereográfica. El plano es tangente y el punto de fuga se sitúa en las antípodas del punto de tangencia. La proyección polar estereográfica es empleada habitualmente para cartografiar las regiones polares.
- Ortográfica. El punto de fuga se sitúa en el infinito.

Existen proyecciones azimutales que no son de tipo perspectivo, es decir, que no se basan en el esquema de la figura 3.5. La proyección de Airy, por ejemplo, es una de ellas.

- Algunas proyecciones **no se ajustan exactamente al esquema planteado**, y no utilizan una superficie desarrollable como tal sino modificaciones a esta idea. Por ejemplo, las proyecciones *policónicas* utilizan la misma filosofía que las cónicas, empleando conos, pero en lugar de ser este único, se usan varios conos, cada uno de los cuales se aplica a una franja concreta de la zona proyectada. La unión de todas esas franjas, cada una de ellas proyectada de forma distinta (aunque siempre con una proyección cónica), forma el resultado de la proyección.
Del mismo modo, encontramos proyecciones como la proyección *sinusoidal*, una proyección de tipo pseudocilíndrico, o la proyección de Werner, cuya superficie desarrollable tiene forma de corazón. Estas proyecciones son, no obstante, de uso menos habitual, y surgen en algunos casos como respuesta a una necesidad cartográfica concreta.

Otra forma distinta de clasificar las proyecciones es según las propiedades métricas que conserven. Toda proyección implica alguna distorsión (denominada *anamorfosis*), y según cómo sea esta y a qué propiedad métrica afecte o no, podemos definir los siguientes tipos de proyecciones:

- **Equiárea**. En este tipo de proyecciones se mantiene una escala constante. Es decir, la relación entre un área terrestre y el área proyectada es la misma independientemente de la localización, con lo que la representación proyectada puede emplearse para comparar superficies.
- **Conformes**. Estas proyecciones mantienen la forma de los objetos, ya que no provocan distorsión de los ángulos. Los meridianos y los paralelos se cortan en la proyección en ángulo recto, igual que sucede en la realidad. Su principal desventaja es que introducen una gran distorsión en el tamaño, y objetos que aparecen proyectados con un tamaño mucho mayor que otros pueden ser en la realidad mucho menores que estos.
- **Equidistantes**. En estas proyecciones se mantienen la proporcionalidad de las distancias.

En los ejemplos de proyecciones que se han citado para los distintos tipos de proyecciones (cónicas, cilíndricas, etc.), puede verse cómo resulta común especificar el tipo en función de la propiedad métrica preservada, para así caracterizar completamente la proyección.

La elección de una u otra proyección es función de las necesidades particulares. Como ya se ha dicho, la proyección polar estereográfica es empleada cuando se trabaja las regiones polares, ya que en este caso es la más adecuada. Proyecciones como la de Mercator, empleadas habitualmente, no resultan tan adecuadas en esas zonas. Asimismo, hay proyecciones que no pueden recoger todo el globo, sino solo una parte de este, por lo que no son de aplicación para pequeñas escalas. La existencia de un gran número de distintas proyecciones es precisamente fruto de las diferentes necesidades que aparecen a la hora de trabajar con cartografía.

3.3.3. El sistema UTM

De entre los cientos de proyecciones que existen actualmente, algunas tienen un uso más extendido, bien sea por su adopción de forma estandarizada o sus propias características. Estas proyecciones, que se emplean con más frecuencia para la creación de cartografía, son también las que más habitualmente vamos a encontrar en los datos que empleemos con un SIG, y es por tanto de interés conocerlas un poco más en detalle.

En la actualidad, una de las proyecciones más extendidas en todos los ámbitos es la proyección universal transversa de Mercator, la cual da lugar al sistema de coordenadas UTM. Este sistema, desarrollado por el ejército de los Estados Unidos, no es simplemente una proyección, sino que se trata de un sistema completo para cartografiar la practica totalidad de la Tierra. Para ello, esta se divide en una serie de zonas rectangulares mediante una cuadricula y se aplica una proyección y unos parámetros geodésicos concretos a cada una de dichas zonas. Aunque en la actualidad se emplea un único elipsoide (WGS–84), originalmente este no era único para todas las zonas.

Con el sistema UTM, las coordenadas de un punto no se expresan como coordenadas terrestres absolutas, sino mediante la zona correspondiente y las coordenadas relativas a la zona UTM en la que nos encontremos.

La cuadricula UTM tiene un total de 60 husos numerados entre 1 y 60, cada uno de los cuales abarca una amplitud de 6°de longitud. El huso 1 se sitúa entre los 180°y 174°O, y la numeración avanza hacia el Este.

En latitud, cada huso se divide en 20 zonas, que van desde los 80°S hasta los 84°N. Estas se codifican con letras desde la C a la X, no utilizándose las letras I y O por su similitud con los dígitos 1 y 0. Cada zona abarca 8 grados de longitud, excepto la X que se prolonga unos 4 grados adicionales.

Una zona UTM se localiza, por tanto, con un número y una letra, y es en función de la zona como posteriormente se dan las coordenadas que localizan un punto. Estas coordenadas se expresan en metros y expresan la distancia entre el punto y el origen de la zona UTM en concreto. El origen de la zona se sitúa en el punto de corte entre el meridiano central de la zona y el ecuador. Por ejemplo, para las zonas UTM en el huso 31, el cual va desde los 0°hasta los 6°, el origen se sitúa en el punto de corte entre el ecuador y el meridiano de 3°(Figura 3.7).

Para evitar la aparición de números negativos, se considera que el origen no tiene una coordenada X de 0 metros, sino de 500000. Con ello se evita que las zonas al Este del meridiano central tengan coordenadas negativas, ya que ninguna zona tiene un ancho mayor de 1000000 metros (el ancho es máximo en las zonas cerca del ecuador, siendo de alrededor de 668 kilómetros).

De igual modo, cuando se trabaja en el hemisferio sur (donde las coordenadas Y serían siempre negativas), se considera que el origen tiene una coordenada Y de 10000000 metros, lo cual hace que todas las coordenadas referidas a él sean positivas.

La figura 3.8 muestra un esquema de la cuadrícula UTM.

Para las zonas polares no resulta adecuado emplear el sistema UTM, ya que las distorsiones que produce son demasiado grandes. En su lugar, se utiliza el sistema UPS (Universal Polar Stereographic).

3.3.4. Transformación y conversión de coordenadas

Una situación muy habitual en el trabajo con un SIG es disponer de cartografía en varios sistemas de coordenadas, o bien en un mismo sistema pero con parámetros diferentes (por ejemplo,

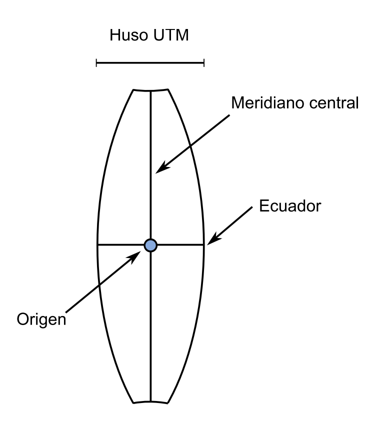


Figura 3.7: Determinación del origen de una zona UTM

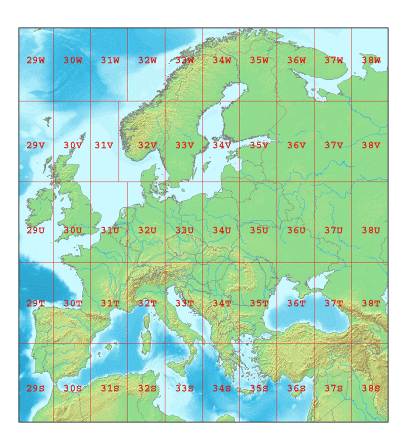


Figura 3.8: Representación parcial de la cuadrícula UTM en Europa (tomado de Wikipedia)

diferente datum). Para poder emplear toda esa cartografía de forma conjunta, resulta necesario trabajar en un sistema único y bien definido, lo cual hace necesario convertir al menos una parte de ella.

Este cambio de coordenadas puede ser obligatorio a cualquier escala de trabajo, ya que las diferencias en el sistema escogido pueden aparecer por circunstancias muy diversas, incluso si todos los datos tienen un origen común. Así, al reunir información de varios países para crear en un SIG un mapa de todo un continente, es probable que los datos de cada país estén referidos a un sistema distinto, pero incluso trabajando en un área más reducida podemos encontrar una situación similar. En España, por ejemplo, podemos encontrar cartografía de algunas Comunidades Autónomas en dos husos UTM distintos, ya que la frontera entre estos cruza y divide dichas Comunidades.

Distinguimos dos tipos de operaciones a realizar con coordenadas:

- **Conversión de coordenadas**. Los sistemas de origen y destino comparten el mismo datum. Es una transformación exacta y se basa en la aplicación de formulas establecidas que relacionan ambos sistemas.
- **Transformación de coordenadas**. El datum es distinto en los sistemas de origen y destino.

Las proyecciones cartográficas, vistas en un punto anterior, son una forma particular de conversión de coordenadas.

Un SIG ha de estar preparado para trabajar con cartografía en cualquiera de los sistemas de referencia más habituales y, más aún, para facilitar al usuario la utilización de todo tipo de información geográfica con independencia del sistema de coordenadas que se emplee. Para ello, los SIG incorporan los procesos necesarios para efectuar cambios de coordenadas, de forma que para unos datos de partida se genera un nuevo conjunto de datos con la misma información pero expresada en un sistema de coordenadas distinto.

Otra forma en la que los SIG pueden implementar estas operaciones es mediante capacidades de transformación y conversión «al vuelo», es decir, en tiempo real. De este modo, pueden introducirse en un SIG datos en sistemas de coordenadas variados, y el SIG se encarga de cambiar estos a un sistema de referencia base fijado de antemano. Este proceso tiene lugar de forma trans-

parente para el usuario, que tiene la sensación de que todos los datos estaban originalmente en el sistema de trabajo escogido.

Esto exige, lógicamente, que todo dato geográfico se acompañe de información acerca del sistema de coordenadas que se ha utilizado para crearlo, algo que no siempre sucede. Veremos más acerca de la importancia de este tipo de información adicional en el capítulo 9.

3.3.5. Codificación de sistemas de referencia

Debido al elevado número de distintos sistemas de referencia existentes, resulta fácil perderse en ellos a la hora de tener que trabajar con cartografía en distintos sistemas. Si bien es cierto que existe un esfuerzo integrador para tratar de homogeneizar el uso de sistemas de referencia, también existen esfuerzos para intentar facilitar la gestión de estos y que no resulte tan complejo combinar cartografía producida utilizando sistemas de coordenadas diferentes.

Uno de los intentos más exitosos en este sentido es el desarrollado por el consorcio petrolífero European Petroleum Survey Group (EPSG), el cual, consciente de la necesidad de disponer de información acerca de los distintos sistemas de coordenadas y de que esta información fuera de fácil acceso y manejo, ha elaborado un esquema de codificación específico.

Este esquema asocia a cada sistema de coordenadas un código (conocido como *código EPSG*) que la identifica. Paralelamente, se han documentado en un formato común las características principales de todos estos sistemas, así como las formulaciones que permiten transformar coordenadas entre ellos.

Esta información constituye el *EPSG geodetic parameter dataset*, un repositorio de los parámetros necesarios para

- identificar coordenadas de tal modo que estas describan la posición de un punto de forma inequívoca y no ambigua.
- definir transformaciones y conversiones que permitan pasar de un sistema de referencia a otro.

3.4. Escala

El concepto de escala es fundamental a la hora de trabajar con cartografía, y es uno de los valores básicos que definen toda representación cartográfica. Esta representación ha de tener un tamaño final manejable, con objeto de que pueda resultar de utilidad y permitir un uso práctico, pero el objeto que se cartografía (un país, un continente o bien la Tierra al completo) es un objeto de gran tamaño. Esto hace necesario que, para crear un mapa, se deba reducir o bien el objeto original o bien el objeto ya proyectado, dando lugar a una versión «reducida» que ya cumple con los requisitos de tamaño adecuado.

Es decir, imaginemos que aplicamos una proyección cónica sobre el elipsoide, empleando para ello un cono que cubra dicho elipsoide. Este cono tendrá que ser, lógicamente, de gran tamaño. Al desarrollarlo, el plano que obtenemos tendrá miles de kilómetros de lado. Debemos fabricar una versión «a escala» de este, que será la que ya podamos utilizar.

En este contexto, la escala no es sino la relación de tamaño existente entre ese gran mapa que se obtiene al desarrollar nuestro cono de proyección y el que finalmente manejamos, de tamaño más reducido. Conociendo esta relación podemos ya conocer las verdaderas magnitudes de los elementos que vemos en el mapa, ya que podemos convertir las medidas hechas sobre el mapa en medidas reales. Es importante recordar que esas medidas no son tan «reales», puesto que la propia proyección las ha distorsionado —lo cual no

debe olvidarse—, pero sí que son medidas en la escala original del objeto cartografiado.

La escala se expresa habitualmente como un denominador que relaciona una distancia medida en un mapa y la distancia que esta medida representa en la realidad. Por ejemplo, una escala 1:50000 quiere decir que 1 centímetro en un mapa equivale a 50000 centímetros en la realidad, es decir, a 500 metros. Conociendo este valor de la escala podemos aplicar sencillas reglas de tres para calcular la distancia entre dos puntos o la longitud de un elemento dado, sin más que medirlo sobre el mapa y después convertir el resultado obtenido en una medida real.

Una vez más es preciso insistir que lo anterior es posible siempre bajo las limitaciones que la propia proyección empleada para crear el mapa tenga al respecto, y que dependerán del tipo de proyección que sea en función de las propiedades métricas que conserva.

De hecho, e independientemente del tipo de proyección, la escala es completamente cierta únicamente en determinadas partes del mapa. Cuando decimos que un mapa tiene una escala 1:50000, este valor, denominado *Escala Numérica*, se cumple con exactitud tan solo en algunos puntos o líneas. En otros puntos la escala varía. La relación entre la escala en esos puntos y la Escala Numérica se conoce como *Factor de Escala*.

A pesar de que la escala es imprescindible para darle un uso práctico a todo mapa, y cualquier usuario de este debe conocer y aplicar el concepto de escala de forma precisa, los SIG pueden resultar engañosos al respecto. Aunque la escala como idea sigue siendo igual de fundamental cuando trabajamos con información geográfica en un SIG, las propias características de este y la forma en la que dicha información se incorpora en el SIG pueden hacer que no se perciba la escala como un concepto tan relevante a la hora de desarrollar actividad con él.

Esto es debido principalmente a que la escala tiene una relación directa con la visualización, ya que se establece entre la realidad y una representación visual particular, esto es, el mapa. Como ya se ha mencionado en el capítulo 1, los datos en un SIG tienen carácter numérico y no visual, y la representación de estos se encarga de realizarla el subsistema correspondiente a partir de dichos datos numéricos. Es decir, que en cierta medida en un SIG no es estrictamente necesaria la visualización de los datos, y cuando esta se lleva a cabo no tiene unas características fijas, ya que, como veremos, el usuario puede elegir el tamaño con el que estos datos se representan en la pantalla.

Un mapa impreso puede ampliarse o reducirse mediante medios fotomecánicos. Sin embargo, no es esta una operación «natural», y está claro que desde el punto de vista del rigor cartográfico no es correcta si lo que se hace es aumentar el tamaño del mapa. En un SIG, sin embargo, es una operación más el elegir la escala a la que se representan los datos y modificar el tamaño de representación, y esta resulta por completo natural e incluso trivial[20].

Pese a ello, los datos tienen una escala inherente, ya que esta no está en función de la representación, sino del detalle con que han sido tomados, y esta escala debe igualmente conocerse para dar un uso adecuado a dichos datos. En este sentido es más conveniente entender la escala como un elemento relacionado con la resolución de los datos, es decir, con el tamaño mínimo cartografiado.

Esta concepción no es en absoluto propia de los SIG, ya que deriva de las representaciones clásicas y los mapas impresos. Se sabe que el tamaño mínimo que el ojo humano es capaz de diferenciar es del orden de 0,2 mm. Aplicando a este valor la escala a la que queremos crear un mapa, tendremos la mínima distancia sobre el terreno que debe medirse. Por ejemplo, para el

caso de un mapa 1:50000, tenemos que la mínima distancia es de 10 metros.

Si medimos puntos a una distancia menor que la anterior y después los representamos en un mapa a escala 1:50000, esos puntos no serán distinguibles para el usuario de ese mapa, y la información recogida se perderá. Estos razonamientos sirven para calcular la intensidad del trabajo que ha de realizarse para tomar los datos con los que después elaborar una determinada cartografía.

En realidad, el concepto de escala no es único, sino que tiene múltiples facetas. Por una parte la escala *cartográfica*, que es la mera relación entre el tamaño en el mapa y la realidad. Por otra, la escala *de análisis* u *operacional*[21], que es la que define la utilidad de los datos y lo que podemos hacer con ellos, ya que indica las limitaciones de estos. Cuando en un SIG aumentamos el tamaño en pantalla de una cierta información geográfica, estamos variando la escala cartográfica, pero no estamos modificando la escala de análisis. Por ello, por mucho que ampliemos, no vamos a ver más detalles, ya que para ello sería necesario tomar más datos.

Veremos más ideas sobre la escala de análisis y algunas implicaciones al respecto en el capítulo 10, al inicio de la parte dedicada a los procesos, ya que estos conceptos son fundamentales para realizar correctamente análisis y operaciones como las descritas en esa parte del libro.

Un tipo de datos espaciales particulares con los que se trabaja en un SIG, los datos *ráster*, tienen a su vez un parámetro de resolución, con una clara relación con el concepto de escala. Veremos más al respecto en el capítulo 5.

3.5. Generalización cartográfica

Muy relacionado con el concepto de escala encontramos la denominada *generalización cartográfica*. Generalizar implicar expresar alguna idea o información de forma más resumida, de tal modo que esta sea comprensible y pueda aprovecharse de la mejor manera posible. En el ámbito cartográfico, cuando hablamos de generalización nos referimos a la adaptación de los elementos de un mapa para lograr que este sea más expresivo y claro, lo cual puede conllevar una distorsión intencional de dichos elementos.

Para ver un ejemplo de lo anterior, pensemos en un mapa del mundo en el que se representen todas las calles y caminos existentes. Esta información tiene una escala adecuada para ser mostrada en un callejero local cuya escala nominal suele ser del orden de 1:5000, pero a la escala 1:1000000, adecuada para un mapa mundial, representar todo su detalle resulta innecesario. La representación resultante va a tener una densidad excesiva, y muchos de sus elementos no podrán distinguirse debido a su cercanía.

En caso de que esta representación no se haga sobre papel sino sobre una pantalla y trabajando con un SIG, la situación es similar y resulta incluso más necesario aplicar alguna forma de generalización. A las limitaciones de la visión humana han de sumarse las limitaciones de resolución que el propio dispositivo presenta. En la situación del ejemplo anterior, muchos elementos del mapa (calles, edificios, etc.), ocuparían por su tamaño un mismo y único punto en la pantalla (veremos más adelante que cada uno de estos puntos se conoce como *píxel*), por lo que resultaría imposible distinguirlos o detallarlos más allá de ese nivel de resolución.

A lo anterior debemos añadir el hecho de que producir esa representación, aunque sea sobre

un solo píxel, puede requerir gran cantidad de procesos y operaciones, ya que el conjunto de calles que se contienen en él pueden presentar gran complejidad, tanto mayor cuanto mayor sea el nivel de detalle con que han sido recogidas en los datos. Es decir, que en el trabajo con un SIG la generalización no tiene importancia únicamente para la visualización en sí, sino también para el rendimiento del propio SIG a la hora de producir dicha visualización.

La generalización, en este caso, implicaría la elección de un subconjunto de todas esas carreteras, de manera que no se representase la totalidad de la información de la que se dispone.

En ocasiones, el proceso de generalización es necesario por razones distintas a lo visto en el ejemplo anterior, y requiere diferentes operaciones. Por ejemplo, podemos crear un mapa del mundo que contenga vías de comunicación, pero no todas, sino solo las principales autopistas de cada país. En este caso, no vamos a encontrar problemas con distintas carreteras que se solapan en la representación, ni tampoco un volumen excesivo de datos, pero debemos igualmente «adaptar» la representación a la escala, es decir, efectuar algún tipo de generalización.

Si en ese mapa representamos una carretera con un ancho de 20 metros a escala 1:1000000, el tamaño que tendrá en el mapa será de tan solo 0,02 milímetros. Este ancho es prácticamente nulo, y no tiene sentido representar esa carretera de esta forma. Ha de representarse con un ancho mayor. Aunque no se esté dibujando con exactitud la magnitud real de ese elemento, el resultado es mucho mejor desde todos los puntos de vista. Esta es otra forma de generalización que busca también mejorar la calidad de la representación y la transmisión de la información que contiene.

La generalización, por tanto, es un proceso que tiene como objetivo la producción de una imagen cartográfica legible y expresiva, reduciendo el contenido del mapa a aquello que sea posible y necesario representar. Para ello, se enfatiza aquello que resulta de importancia y se suprime lo que carece de ella [22].

3.5.1. Operaciones de generalización

Existen diversas operaciones que se emplean en el proceso de generalización. Algunas de las más relevantes son las siguientes [23]:

- **Simplificación**. Se trata de crear elementos más sencillos que sean más fáciles y rápidos de representar. Los elementos originales se sustituyen por estos más sencillos, de tal modo que se mantienen las características visuales principales pero las operaciones con los datos se optimizan.
- **Suavizado**. Se sustituyen formas angulosas por otras más suaves y de menor complejidad.
- **Agregación**. Un conjunto de varios objetos se sustituye por uno nuevo con un menor número. Por ejemplo, al representar una ciudad, no dibujar cada una de las casas, sino solo el contorno de cada manzana. La figura 3.9 muestra un ejemplo de esta técnica aplicado a elementos lineales, en particular carreteras.
- **Exageración**. En ocasiones, mantener el objeto a la escala que le corresponde haría que no se pudieran apreciar las características de este. En este caso, se exagera su tamaño para que pueda interpretarse con mayor facilidad y no perder información en la representación.
- **Desplazamiento**. Un objeto se representa en una posición distinta a la que le corresponde, con el fin de garantizar su visibilidad y obtener un resultado más claro.

Combinando operaciones como las anteriores de forma adecuada, se obtiene una cartografía mucho más útil y con mayor potencial. En el caso de trabajar en un SIG, algunas de estas operaciones, como pueden ser la simplificación o la

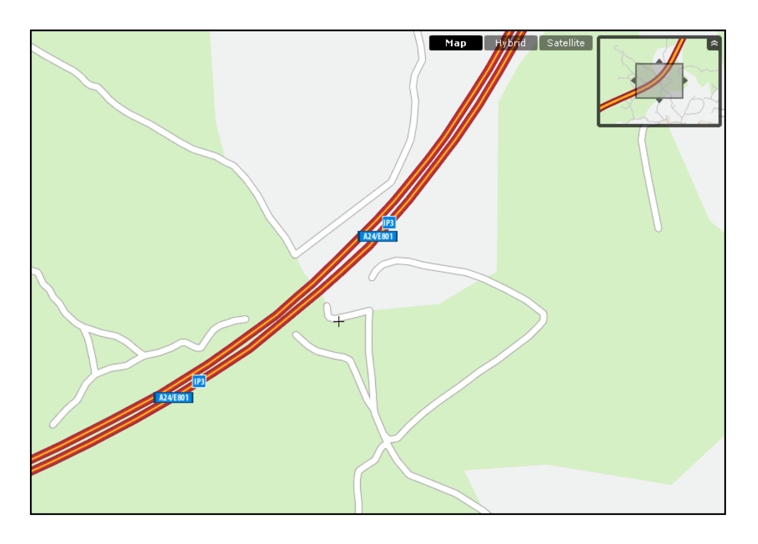


Figura 3.9: Un ejemplo de generalización por agregación. Dos carreteras prácticamente paralelas y unidas se representan como dos elementos en el mapa, pero en el localizador de la parte superior izquierda, a escala de menor detalle, se generalizan como una única (Tomado de Yahoo Maps).

agregación, tiene también un efecto beneficioso sobre el propio manejo de los datos dentro del SIG.

Estas operaciones se enumeran aquí como ideas a aplicar para efectuar la generalización de un documento geográfico, como corresponde a este capítulo de fundamentos y conceptos cartográficos básicos. No obstante, también las veremos en otras partes del libro, ya que no son exclusivas de esta parte. Por su importante papel en la representación visual de los datos, veremos más al respecto en la parte dedicada a visualización. Algunos algoritmos para la simplificación y suavizado de líneas los estudiaremos en la parte dedicada a procesos, particularmente en el apartado 19.7.

3.5.2. Generalización en el contexto de un SIG

La generalización es importante en un SIG debido a la variedad de escalas posibles que puede tener la información con que se trabaja, así como por la variedad de escalas de representación que pueden definirse gracias a la flexibilidad que el propio SIG presenta en sus capacidades de visualización. Existen diversas formas de enfocar inicialmente el problema de obtener un juego de datos óptimo para ser representado en cada caso y una representación óptima de este.

La mayor problemática se encuentra en el manejo de datos con gran precisión y gran volumen —como, por ejemplo, esos datos de calles y vías de todo el mundo— al representarlos a una escala de menor detalle, aunque el proceso de generalización no es necesario exclusivamente en este caso, sino en muchos otros con independencia del volumen y la escala original.

Una aproximación básica puede ser trabajar con todo el conjunto de datos y generalizarlo a medida que sea necesario en función de la escala de trabajo en cada momento. Es decir, si el usuario decide visualizar todo un continente, el SIG no traza todas las calles de ese continente, sino que se seleccionan de forma automática los objetos a representar y después se crea la repre-

sentación. Las operaciones de generalización se llevan a cabo en el momento mismo en que el usuario lo necesita.

Este tipo de generalización «al vuelo» no resulta, sin embargo, óptimo, y en la mayoría de los casos es inviable o no proporciona los resultados esperados. Esto es así debido a que se ha de trabajar con el gran volumen de datos original, y generalizar estos es una tarea suficientemente compleja como para que los algoritmos encargados de hacerlo no lo hagan de forma fluida. No ha de olvidarse que, mientras que la razón fundamental de la generalización en el contexto de la cartografía clásica es la mera visualización y la transmisión de la información, en el entorno de un SIG también existen razones relacionadas con la eficiencia de los procesos, como ya se ha mencionado. Aplicando esta metodología, la generalización no es ventajosa en términos de cómputo, sino que, por el contrario, puede incluso suponer una carga adicional al proceso de visualización.

Aun en el caso de que el volumen de datos no fuera grande y no existieran problemas de rendimiento, una generalización por completo automatizada no garantiza un resultado óptimo. Pese a que existen algoritmos y formulaciones matemáticas que permiten generalizar de forma relativamente adecuada (algunos de los cuales los veremos más adelante en este libro), el proceso global de generalización combina varios procedimientos distintos, y en conjunto conforma un proceso no exento de subjetividad. La labor tradicional del cartógrafo no puede automatizarse de forma total, y se hace necesario cierto trabajo manual para obtener un resultado de calidad o evaluar el generado por un procedimiento automático.

Por todo lo anterior, la forma de incorporar la generalización dentro de un SIG suele basarse en un enfoque multi–escalar, en el cual se maneja información de una misma zona de estudio a diferentes escalas, y se usa en cada momento aquella que resulte más conveniente. Si trabajara con cartografía en papel, sería equivalente a tener varios mapas de una zona a diferentes escalas.

Por ejemplo, en un mapa con núcleos de población a escala 1:25000 se almacenará cada ciudad como un polígono que delimite su contorno. Esa misma información a escala 1:1000000 se almacenará como un único punto cada ciudad, ya que el tamaño de esta es demasiado pequeño en la representación, y no tiene sentido el empleo de tanto detalle. Para convertir un mapa en otro se ha producido un proceso de simplificación, convirtiendo polígonos en puntos.

Si incorporamos ambos mapas dentro de un SIG, podemos utilizar el que corresponda en función de la escala requerida. De este modo, la generalización no es una tarea que el propio SIG desarrolle, sino que cuando esta es necesaria puede recurrir a una información ya generalizada de antemano. El rendimiento del proceso es mayor, y además el dato generalizado puede haber sido elaborado de la forma más conveniente.

El concepto de *capa*, que veremos en el capítulo 4 y que es vital para la idea actual de un SIG, permite este manejo simultáneo de información a distintas escalas.

En la figura 3.10 puede verse un esquema de lo anterior. A medida que variamos la escala de representación, la información que vemos representada tiene una escala distinta y podría también tener un distinto origen. Incluso el tipo de información que vemos varía, ya que las representaciones más globales son de tipo gráfico, creadas a partir de los propios datos almacenados como objetos (calles, carreteras, etc.), mientras que la de mayor detalle es una fotografía aérea.

En el caso de imágenes tales como esa fotografía aérea, existen además en un SIG una serie de procesos que también pueden considerarse como parte de la generalización, y que atañen más al

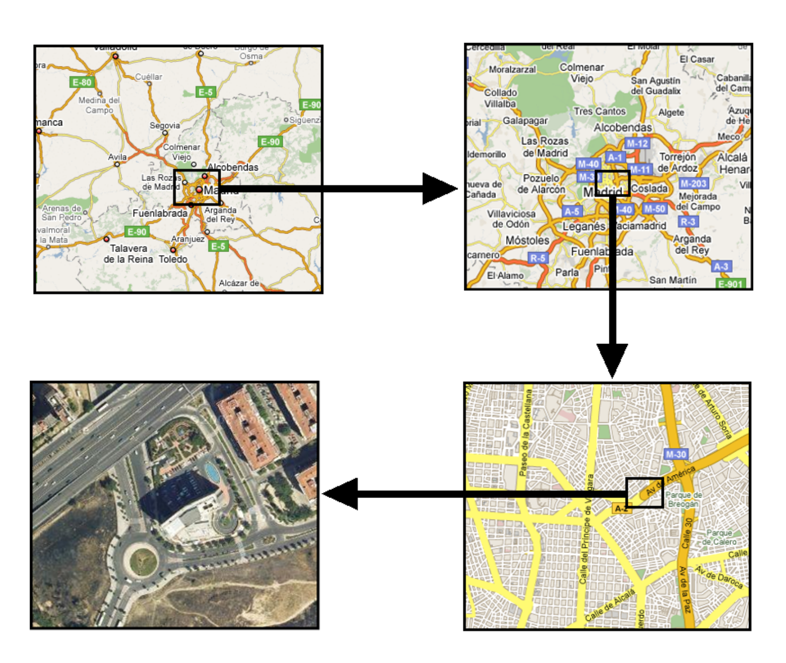


Figura 3.10: En un SIG es habitual manejar información a diferentes escalas. En función de la escala de representación, la información visualizada será una u otra.

rendimiento que a la representación. Para entenderse esto piénsese que las imágenes se componen de elementos denominados *píxeles*, que son pequeños puntos, cada uno de los cuales tendrá un color asociado (esto lo veremos con mucho más detalle en el capítulo 5). El numero de estos píxeles en una imagen grande es muy superior al de una pantalla (una pantalla también se divide en puntos, si te acercas a una lo podrás ver claramente).

El proceso de representación de la imagen en la pantalla consiste en calcular qué color asignar a cada píxel de la pantalla en función de los de la imagen, pero este proceso, si se utiliza la imagen completa, es muy costoso en términos de cómputo, ya que implica procesar toda la información de la imagen, que puede ser del orden de centenares de millones de píxeles. Si representamos una porción de esa imagen (una porción del territorio que cubre), podemos solo trabajar con los píxeles en esa zona, pero la representación de toda la imagen hace necesario procesar todos los valores que contiene.

Este proceso en realidad puede verse como un tipo de generalización «al vuelo». Ya dijimos que este tenía principalmente dos problemas: el rendimiento y la imposibilidad de obtener resultados óptimos de forma automatizada. En el caso de imágenes, existe el problema del rendimiento, pero es posible automatizar la creación de datos a diferente escala de trabajo. Esto es así debido a que la representación de elementos tales como carreteras o lagos se hace mediante una interpretación de esos objetos, y este proceso es en cierta medida subjetivo, como vimos. En el caso de imágenes no hay que interpretar objeto alguno, ya que esos objetos ya «están representados» en la imagen, y únicamente es necesario disminuir la escala.

Los algoritmos para llevar a cabo este proceso se conocen como de *remuestreo*, y los veremos con detalle en el capítulo 15. Algunos SIG utilizan estos algoritmos para hacer más fluido el manejo de grandes imágenes mediante la creación de las denominadas *pirámides*. Cuando el usuario introduce en el SIG una imagen de gran tamaño, este prepara varias versiones de esa imagen a distintas escalas de detalle, de forma que posteriormente pueda recurrir a la que sea más conveniente en cada caso en función de la escala de representación. Es decir, el SIG realiza la «generalización» de esa imagen de forma automática, siendo necesario proporcionarle únicamente la imagen de mayor detalle. La figura 3.11 ilustra gráficamente esto.

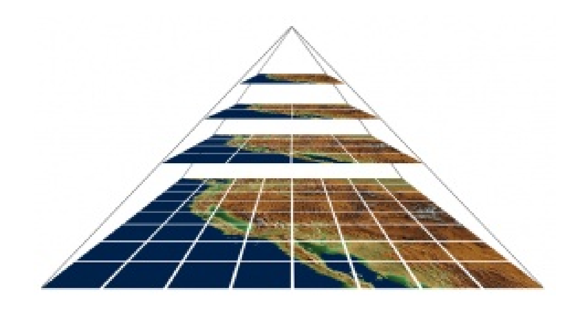


Figura 3.11: Pirámides de representación con imágenes preparadas a distintas escalas (Fuente: OSGeo).

3.6. Resumen

La cartografía y la geodesia son ciencias que aportan un importante conjunto de conocimientos y elementos al mundo de los SIG, y su estudio es fundamental para cualquier trabajo con un SIG.

La geodesia se encarga de estudiar la forma de la Tierra, con objeto de posteriormente poder localizar de forma precisa los puntos sobre esta mediante un sistema de coordenadas. Dos conceptos básicos en geodesia son el *geoide* y el *elipsoide*, superficies de referencia que modelizan la forma de la Tierra. El primero es la superficie formada por los puntos en los que el campo gravitatorio tiene una misma intensidad, y se obtiene prolongando la superficie de los océanos en reposo bajo la superficie terrestre. El segundo es un objeto definido por una ecuación y una serie de parámetros, que permite asimilar la Tierra a una superficie matemática.

El conjunto de un elipsoide y un punto de tangencia con la superficie terrestre (Punto Fundamental), forma un *datum*.

Para asignar coordenadas a un punto en función de los elementos anteriores es necesario definir un sistema de referencia. Las coordenadas geográficas han sido utilizadas tradicionalmente, y son de utilidad para grandes zonas. Otro tipo de coordenadas más intuitivas son las cartesianas, y para su obtención se requiere el concurso de una proyección cartográfica que convierta coordenadas espaciales en coordenadas planas. Hay muchos tipos de proyecciones, siendo el sistema UTM uno de los más extendidos.

En el ámbito de la cartografía, hemos visto en este capítulo la importancia del concepto de escala, que no pierde su papel fundamental al trabajar en un SIG en lugar de hacerlo con cartografía impresa. Estrechamente relacionada con la escala encontramos la *generalización*, que comprende una serie de procesos encaminados a la obtención de una representación lo más clara posible de una serie de datos a una escala dada.

Parte II

Los datos

Capítulo 4

¿Con qué trabajo en un SIG?

4.1. Introducción

De todos los subsistemas de SIG, el correspondiente a los datos es el pilar fundamental que pone en marcha los restantes. Los datos son el combustible que alimenta a los restantes subsistemas, y sin los cuales un SIG carece por completo de sentido y utilidad.

El subsistema de datos es, a su vez, el más interrelacionado, y está conectado de forma inseparable a todos los restantes. Mientras que, por ejemplo, la visualización no es imprescindible para el desarrollo de procesos de análisis, no hay elemento del sistema SIG que pueda vivir si no es alimentado por datos. Los datos son necesarios para la visualización, para el análisis y para dar sentido a la tecnología y, en lo referente al factor organizativo y a las personas, el rol de estas en el sistema SIG es en gran medida gestionar esos datos y tratar de sacar de ellos el mayor provecho posible. Por tanto, los datos son fundamentales en un SIG, y todo esfuerzo dedicado a su estudio y a su mejor manejo será siempre positivo dentro de cualquier trabajo con SIG.

La forma en que los datos se gestionan en un SIG es un elemento vital para definir la propia naturaleza de este, así como sus prestaciones, limitaciones y características generales.

4.2. Datos *vs* Información

Existe una importante diferencia entre los conceptos de *datos* e *información*. Ambos términos aparecen con frecuencia y pueden confundirse, pese a que representan cosas bien diferentes. Un SIG es un Sistema de *Información* Geográfica, pero maneja *datos* geográficos, existiendo diferencias entre estos conceptos.

Entendemos como dato al simple conjunto de valores o elementos que utilizamos para representar algo. Por ejemplo, el código 502132N es un dato. Este código por sí mismo no tiene un significado, y es necesario interpretarlo para que surja ese significado. Al realizar esa interpretación, el dato nos *informa* del significado que tiene, y es en ese momento cuando podemos emplearlo para algún fin y llevar a cabo operaciones sobre él que tengan sentido y resulten coherentes con el significado propio que contiene.

El dato anterior podemos interpretarlo como si fuera una referencia geográfica, y cuyo significado sería entonces una latitud, en particular $50^\circ 21' \, 32''$ Norte. Si lo interpretamos como un código que hace referencia a un documento de identificación de una persona, la información que nos aporta es en ese caso completamente distinta. El dato sería el mismo, formado por seis dígitos y una letra, pero la información que da es diferente, ya que lo entendemos e interpretamos de manera distinta.

La información es, por tanto, el resultado de un dato y una interpretación, y el trabajo con datos es en muchos casos un proceso enfocado a obtener de estos toda la información posible. Un dato puede esconder más información que la que a primera vista puede apreciarse, y es a través de la interpretación de los datos como se obtiene esta.

En el capítulo 16 veremos cómo a partir de un Modelo Digital de Elevaciones podemos calcular parámetros tales como la pendiente, extraer el trazado de la red de drenaje o delimitar las subcuencas en que una cuenca vertiente mayor puede dividirse. El dato en este caso lo constituyen los valores que representan la elevación en los distintos puntos. La información que contienen está formada por todo ese conjunto de elementos que podemos obtener, desde la pendiente a los cursos de los ríos, pasando por todo aquello que mediante la aplicación de procesos u operaciones de análisis podamos extraer de esos datos.

Comprender el significado y las diferencias entre datos e información permiten entender, entre otras cosas, que la relación entre los volúmenes de ambos no es necesariamente constante. Por ejemplo, los datos 502132NORTE o CINCUENTA VEINTIUNO TREINTAYDOS NORTE son mayores en volumen que 502132N, pero recogen la misma información espacial que este (suponiendo que los interpretamos como datos de latitud). Tenemos más datos, pero no más información. Podemos establecer planteamientos basados en este hecho que nos ayuden a almacenar nuestra información geográfica con un volumen de datos menor, lo cual resulta ventajoso. Veremos algunos de estos planteamientos más adelante dentro de esta parte del libro.

Aspectos como estos son realmente mucho más complejos, y el estudio de la relación entre datos e información y sus características no es en absoluto sencilla. Existe una disciplina, la *ciencia de la información*, dedicada a estudiar los aspectos teóricos relativos a la información y la forma en que esta puede contenerse en los datos. El lector interesado puede consultar [24, 25] para saber más al respecto.

En este capítulo de introducción a la parte dedicada a los datos, veremos más acerca de la información que de los datos espaciales, pues la manera en que concebimos esta condiciona la forma de los datos. Será en el capítulo siguiente cuando tratemos ya los datos, abordando uno de los problemas fundamentales: la creación del dato espacial.

4.3. Las componentes de la información geográfica

Comprender la información geográfica es vital para poder capturar dicha información e incorporarla a un SIG. En líneas generales, podemos dividir esta en dos componentes principales, cada una de los cuales tiene su implicación particular en los procesos posteriores de representación que más adelante veremos.

- Componente espacial
- Componente temática

La componente espacial hace referencia a la posición dentro de un sistema de referencia establecido. Esta componente es la que hace que la información pueda calificarse como *geográfica*, ya que sin ella no se tiene una localización, y por tanto el marco geográfico no existe. La componente espacial responde a la pregunta *¿dónde?*

Por su parte, la componente temática responde a la pregunta *¿qué?* y va invariablemente unida a la anterior. En la localización establecida por la componente espacial, tiene lugar algún proceso o aparece algún fenómeno dado. La naturaleza de dicho fenómeno y sus características particulares quedan establecidas por la componente temática.

Puede entenderse lo anterior como una variable fundamental (la componente temática), que se sirve de una variable soporte (la componente espacial) para completar su significado.

Los tipos de división horizontal y vertical de la información que veremos más adelante implican una separación en unidades, que en la práctica puede implicar en un SIG que cada una de esas unidades quede almacenada en un lugar o fichero distinto. En el caso de las componentes temática y espacial de la información, son posibles distintos enfoques, ya que estas pueden almacenarse de forma conjunta o bien por separado. El capitulo 8 trata estos enfoques, y en él veremos con detalle cómo puede abordarse el almacenamiento de ambas componentes de la mejor forma posible, así como la evolución que se ha seguido al respecto dentro del campo de los SIG.

Mientras que la componente espacial va a ser generalmente un valor numérico, pues son de esa naturaleza los sistemas de coordenadas que permiten expresar una posición concreta en referencia a un marco dado, la componente temática puede ser de distintos tipos:

- **Numérica**. A su vez, pueden señalarse los siguientes grupos:
  - **Nominal**. El valor numérico representa una identificación. Por ejemplo, el número de un portal en una calle, o el numero del DNI de una persona. Este tipo de variable, al igual que la de tipo alfanumérico, es de tipo cualitativo, frente a las restantes que son de tipo cuantitativo.
  - **Ordinal**. El valor numérico establece un orden. Por ejemplo, una variable que recoja el año de fundación de una serie de ciudades.
  - **Intervalos**. Las diferencias entre valores de la variable tienen un significado. Por ejemplo, entre dos valores de elevación.
  - **Razones**. Las razones entre valores de la variable tienen un significado. Por ejemplo, podemos decir que una precipitación media de 1000mm es el doble que una de 500mm. La pertenencia de una variable a un grupo u otro no solo depende de la propia naturaleza de la misma, sino también del sistema en que se mida. Así, una temperatura en grados centígrados no se encuentra dentro de este grupo (pero sí en el de intervalos), ya que la razón entre dichas temperaturas no vale para decir, por ejemplo, que una zona está al doble de temperatura que otra, mientras que si expresamos la variable temperatura en grados Kelvin sí que podemos realizar tales afirmaciones. El valor mínimo de la escala debe ser cero.
- **Alfanumérica**

El tipo de variable condiciona las operaciones que pueden realizarse con un dato geográfico en función de cómo sea su componente temática. Por ejemplo, carece sentido realizar operaciones aritméticas con variables de tipo ordinal o nominal, mientras que es perfectamente lógico con los restantes tipos dentro de la categoría numérica. También, como veremos en el capítulo 28, influye en la forma de representarlo a la hora de elaborar cartografía.

Además de las componentes espacial y temática, Sinton [26] añade la componente temporal y propone un esquema sistemático que permite clasificar en grupos las distintas clases de información geográfica. Según este esquema, cada una de estas componentes puede estar en uno de los siguientes tres estados posibles: *fija, controlada* o *medida.* Al medir una de estas componentes, es necesario controlar otra de ellas, y fijar la tercera, o bien ignorarla y no tenerla en cuenta (este era el caso explicado hasta el momento, en el cual no habíamos citado aún la componente temporal)

Por ejemplo, si registramos la temperatura a lo largo de un periodo de tiempo para un punto concreto, la componente temporal está controlada (tomamos mediciones de temperatura con un intervalo de tiempo establecido), la componente temática (la propia temperatura) está medida, y la componente espacial está fija (el termómetro que registra los valores se encuentra siempre en un punto inmóvil)

En general, la información geográfica se recoge haciendo fija la componente temporal, y midiendo o controlando las restantes en función del tipo de información de que se trate.

Un concepto a tener en cuenta en relación con las componentes de la información geográfica es la *dimensión*. Los elementos que registramos pueden ir desde sencillos puntos (0D) hasta volúmenes tridimensionales (3D). Un caso particular —y muy frecuente— lo encontramos cuando estudiamos la forma tridimensional del terreno, pero tratando la elevación como variable temática, no como una parte más de la componente espacial. En este caso, tenemos una serie de valores de elevación (Z) localizados en el plano XY. Esto no es realmente equivalente a utilizar una componente espacial tridimensional, ya que no permite recoger en un mismo punto distintos valores (no puede, por ejemplo, modelizarse la forma de una cueva o un objeto vertical), por lo que se conoce como representación en 2.5 dimensiones (2.5D). La figura 4.1 muestra esquemáticamente el concepto de dimensión de los datos dentro de un SIG.

Por ultimo, un aspecto importante de toda variable estudiada es su *continuidad*. Se entiende esta continuidad como la capacidad de la variable para tomar todos los valores dentro de un rango definido. La temperatura, la presión o la elevación son valores continuos, mientras que ninguna variable de tipo nominal puede ser continua, ya que se encuentra limitada a un numero (finito) de identificadores posibles. Por ejemplo, en el caso del número de un DNI, los valores son siempre enteros, existe el valor 1 y el valor 2, pero no los infinitos valores decimales entre ambos.

La continuidad de la variable temática se puede estudiar igualmente en relación con la componente espacial. Así, existen variables que varían de forma continua en el espacio, mientras que otras no lo hacen. Se emplea aquí el concepto matemático de continuidad, es decir, que si trazáramos un perfil de la variable a lo largo de un recorrido dado, la representación de dicho perfil sería una curva que podría dibujarse sin levantar el lápiz del papel.

Todas estas ideas referidas a las distintas variables (distintas informaciones que pretendemos recoger de una zona de estudio dada) nos servirán para detallar los diferentes enfoques de representación y almacenamiento que veremos en el próximo capítulo, y escoger en cada caso el más apropiado.

4.4. División horizontal de la información geográfica

Además de dividir la información geográfica en componentes, también dividimos esta con criterios puramente espaciales, «cortándola» en unidades menores que ocupen una región de amplitud más reducida. Este es un procedimiento similar al que encontramos en un mapa impreso, ya que el territorio de un país se encuentra cartografiado en diferentes *hojas*. Las razones para esto son, por una parte, los posibles distintos orígenes que los diferentes mapas pueden tener (cada región puede ser responsable de fabricar los suyos) y, especialmente, el hecho de que, de no ser así, los mapas tendrían un tamaño inmanejable. Si cartografiamos a escala 1:25000 todo un país, es obvio que no podemos hacerlo en un único mapa, ya que este sería enorme.

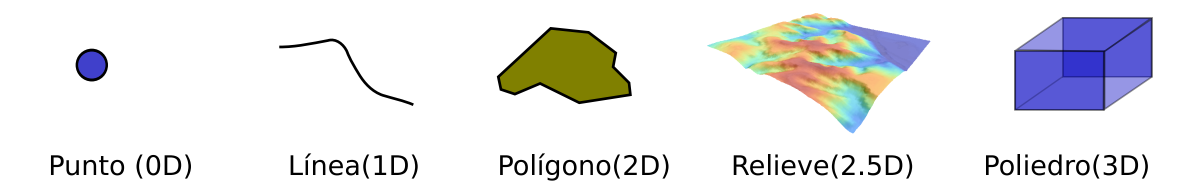


Figura 4.1: Dimensión de los datos geográficos

En el caso de trabajar en un SIG, no tenemos el problema del tamaño físico del mapa, ya que no existe tal tamaño. Los datos no ocupan un espacio físico, pero sí que requieren un volumen de almacenamiento, y este presenta el mismo problema. Recoger a escala 1:25000 todo un país supone un volumen de datos enorme, que es conveniente dividir para poder manejar con fluidez.

En ambos casos, ya sea dentro de un SIG o no, suele resultar necesario emplear varios bloques de información (varias hojas) para cubrir un área de trabajo. En esta circunstancia, las propias características de un SIG y su forma de trabajo con los datos hacen que este proceso sea más sencillo y eficaz.

La principal cualidad de un SIG para integrar de forma transparente datos correspondientes a zonas distintas y formar un mosaico único es la separación que existe entre datos y visualización. Los datos son la base de la visualización, pero en un SIG estos elementos conforman partes del sistema bien diferenciadas. Esto quiere decir que los datos se emplean para crear un resultado visual pero en sí mismos no contienen valores relativos a esa visualización.

De este modo, es posible combinar los datos y después representarlos en su conjunto. Un proceso así no puede realizarse con un mapa ya impreso, pues este contiene ya elementos de visualización e incluso componentes cartográficos tales como una flecha indicando el Norte, una leyenda o una escala. Por ello, aunque puedan combinarse, realmente no se «funde» la información de cada uno de los mapas para conformar uno único. Dicho de otro modo, si tomamos cuatro hojas contiguas de una serie de mapas no podemos formar un nuevo mapa que sea indistinguible de uno cuatro veces más grande que haya sido impreso en un único pliego de papel.

En un SIG, sin embargo, sí que sucede así, y la visualización de cuatro o más bloques de datos puede ser idéntica a la que obtendría si todos esos datos constituyeran un único bloque. Empleando herramientas habituales en un SIG, y si cada uno de esos bloques está almacenado en un fichero, resulta incluso posible unirlos todos y crear un solo fichero que los contenga.

Una de las razones principales que favorecen esta combinación de datos es el hecho de que la escala nominal es en sí un elemento de representación. Como vimos en el apartado 3.4, la escala nominal relaciona el tamaño que tiene un objeto en la representación con su tamaño real, y la forma en que se recoge la información a la hora de realizar medidas de ese objeto viene condicionada por dicha escala, de tal modo que el esfuerzo desarrollado en esas mediciones sea coherente con la representación que se va a hacer posteriormente al crear el mapa.

Los datos que manejamos en un SIG tiene una escala de detalle impuesta por la precisión de las mediciones, pero no una escala nominal asignada, ya que no tienen un tamaño fijo de representación en la pantalla del ordenador o el periférico correspondiente, al contrario que un mapa impreso en el que los distintos elementos

ya se encuentran representados. Esto hace que combinar cartografía clásica a distintas escalas sea complejo, ya que los mapas no «casan» bien entre sí.

En el caso de un SIG, es el usuario el que decide la escala de representación, y esta será la misma para todos los datos que se visualicen, independientemente de las características de estos. En el contexto actual de datos geográficos, es habitual encontrar situaciones en las que para una zona de terreno disponemos de información a una escala, y para otra zona contigua a esta la información disponible es a una escala distinta. Con el uso de un SIG, sin embargo, es posible trabajar sin problemas con todo el conjunto, sin preocuparse por la integración de sus distintas partes.

Lógicamente, no debe dejarse de lado nunca el rigor cartográfico y, como se dijo en su momento, no olvidar que, aunque podamos representar cualquiera de esos datos a la escala que deseemos, los datos en sí no son suficientes para ello y tienen unas limitaciones impuestas por su escala inherente. Es decir, que no es necesario preocuparse por la integración a la ahora de visualizar y gestionar los datos, pero sí a la hora de analizarlos u obtener resultados a partir de ellos. No obstante, el proceso de combinación es en cualquier caso transparente para el usuario que visualiza esos datos en un SIG, y la operación pasa de ser algo tedioso y complejo a algo prácticamente inapreciable dentro del SIG, pues es este quien se encarga de ocultar toda esa complejidad y simplemente generar las representaciones según los parámetros requeridos en cada momento.

La figura 4.2 muestra un ejemplo de lo anterior en el que puede verse cómo varias fotografías aéreas forman un mosaico que cubre una zona dada, teniendo estas distinto nivel de detalle.

4.5. División vertical de la información. Capas

Uno de los grandes éxitos de los SIG es su estructura de manejo de información geográfica, que facilita todas las operaciones que se llevan a cabo con esta. El concepto de *capa*, imprescindible para comprender todo SIG, es una de las grandes virtudes inherentes a los Sistemas de Información Geográfica, ya que favorece la correcta estructuración de la información y el trabajo con ella.

La división horizontal que ya hemos visto no es algo nuevo, y la gran mayoría de los mapas clásicos cubren una porción relativamente pequeña de la superficie terrestre. Combinando distintos mapas podemos formar uno mayor que cubra una extensión más amplia, y aunque ya hemos visto que esto mismo puede realizarse con un SIG y la tarea resulta así más sencilla, no resulta una operación tan compleja y extraña en el caso de no trabajar en un entorno SIG.

Más difícil, sin embargo, es combinar distintos tipos de información, como por ejemplo la contenida en un mapa topográfico y la existente en un mapa de tipos de suelo y otro de vegetación potencial. Para una misma zona, trabajaremos con varios mapas simultáneamente, y combinar estos para la realización de operaciones en las que intervengan todos ellos(supongamos, por ejemplo, calcular el área total de las zonas con un tipo de suelo dado donde la vegetación corresponde a una clase concreta y se encuentran por encima de 1000 metros) es difícil y generalmente también impreciso.

En el caso de un SIG, los distintos tipos de información se pueden combinar de forma sencilla y limpia, y no aparecen los mismos problemas. Esto es así debido a que la idea de capa permite dividir la información espacial referida a una zona de estudio en varios niveles, de tal forma que,



Figura 4.2: Integración de datos en sentido horizontal. A pesar de que la escala de detalle es distinta para las fotografías aéreas de la imagen, estas se combinan sin problema en un SIG, representándose a una escala dada todas ellas de forma sencilla. Nótese la mayor definición en la parte inferior de la imagen, que se forma con imágenes tomadas a una escala distinta a la de las de la parte superior. Adviértase igualmente la distinta iluminación, ya que han sido tomadas en fecha y horas distintas.

pese a coincidir sobre un mismo emplazamiento, información sobre distintas variables se encuentra recogida de forma independiente. Es decir, en función de la componente temática se establecen distintos bloques de datos espaciales.

Para comprender mejor el concepto de capa, pensemos en un mapa topográfico clásico. En él vamos a encontrar elementos como curvas de nivel, carreteras, núcleos urbanos, o simbología relativa a edificios y puntos singulares (iglesias, monumentos, etc.) Todos estos elementos en su conjunto componen el mapa, y aparecen en una misma hoja como una unidad coherente de información geográfica. No obstante, cada uno de los de estos grupos de información recogidos —elevaciones, red viaria, núcleos urbanos, puntos de interés arquitectónico— pueden recogerse de forma independiente, y combinarse al componer el mapa según las necesidades del momento, o bien combinarse de modo distinto o emplearse individualmente (Figura 4.3).

La figura es lo suficientemente gráfica como para entender la razón de que a este tipo de división la denominemos *vertical*, así como el propio nombre de *capa*, ya que de ella resulta una serie de diferentes niveles que se pueden superponer según el criterio particular de cada usuario de SIG.

Toda la información geográfica con que trabajemos en un SIG va a ser en forma de capas. Cada una de estas capas puede abrirse de forma independiente en un SIG y utilizarse por sí misma o en conjunto con otras en la combinación que se desee.

Esta forma de proceder no es exclusiva de los SIG, y antes de la aparición de estos ya existían experiencias previas en este sentido, combinándose capas de información geográfica para la realización de análisis (véase 2.6). Es, sin embargo, con la aparición de los SIG cuando esta metodología se aplica de forma regular y se establece sistemá-

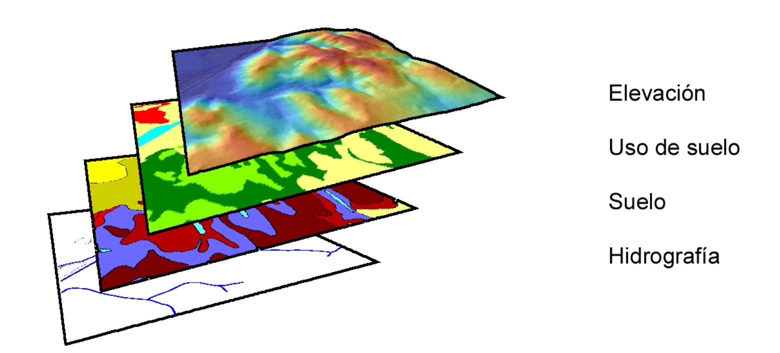


Figura 4.3: Concepto de *capa* de información geográfica dentro de un SIG

ticamente dicha estructuración de la información geográfica.

Así, la visualización, el análisis, y todas las acciones que se realizan sobre la información geográfica dentro de un SIG, se llevan a cabo sobre un conjunto de capas, entendiéndose cada una de ellas como la unidad fundamental de información sobre una zona dada y un tipo de información concreta.

La relevancia del concepto de capa como elemento fundamental de un SIG es enorme, pues realmente constituye el marco básico sobre el que se van a llevar a cabo gran parte de las operaciones. Algunas de las posibilidades que brinda esta filosofía ya las conocemos. Por ejemplo, vimos en el apartado dedicado a la generalización cartográfica cómo en un SIG podemos utilizar diferentes «versiones» de los datos correspondientes a una zona concreta, y representar una u otra de ellas en función de la escala de trabajo. Para un tipo de información, por ejemplo los usos del suelo, estas versiones se almacenarán como distintas capas. La capa es así la unidad fundamental no solo en términos de un área dada, sino también de una escala concreta, y permite una división de los datos óptima a todos los efectos.

Al igual que veíamos en el apartado anterior, las capas nos van a permitir la combinación de datos a distinta escala, no ya en este caso datos contiguos, sino datos correspondientes a un mismo área pero con variables distintas. Esto es de gran utilidad en el trabajo habitual, ya que no todas las variables se recogen con un mismo nivel de detalle, y el detalle con el que podemos encontrar una capa de elevaciones va a ser generalmente mucho mayor que el que cabe esperar para una capa de, digamos, litología.

En realidad, y en el lenguaje habitual de trabajo con SIG, la capa no define únicamente una división vertical, sino también una horizontal. Es más sencillo visualizar la idea de capa con un esquema como el de la figura 4.3, en el que las distintas variables se «apilan» en capas de información superpuestas. Sin embargo, las divisiones horizontales en un mosaico de datos también se consideran como capas distintas en un SIG, pese a contener una misma variable y un mismo tipo de información. Por tanto, y aunque la división vertical sea la que verdaderamente define la idea de capa, cuando hablamos de una capa de datos en un SIG nos referimos a un «trozo» de toda la información disponible, que implica una sección en la dimensión vertical (la de las variables existentes que pueden estudiarse) y un recorte en la horizontal (la de la superficie geográfica).

Las capas pueden emplearse también para incorporar en cierta forma la variable temporal si se considera que la dimensión vertical es el tiempo. Aunque no es la manera más adecuada, y en la actualidad el manejo del tiempo es uno de los principales problemas a resolver en el diseño de los SIG, podemos trabajar con varias capas que representen una misma información y una misma zona, pero en instantes distintos. Esto no es distinto a trabajar con mapas clásicos correspondientes a diferentes instantes, salvo que en el caso de capas cada elemento de la información se encuentra separado a su vez.

Por último, es importante el hecho de que la separación de la información en capas evita la redundancia de datos, ya que cada capa contiene un tipo de información concreto. En un mapa clásico se presentan siempre varias variables, algunas de ellas presentes con carácter general, tales como nombres de ciudades principales o vías más importantes de comunicación. Es decir, que un mapa de usos de suelo o un mapa geológico van a contener otras variables, que en ocasiones se añaden a este para enriquecerlo. Unas curvas de nivel, por ejemplo, permitirán una mejor interpretación de esa geología.

Al dividir toda la información en capas, podemos combinar curvas de nivel y geología, añadir otros elementos, o bien representarlas de forma aislada, algo que no resulta posible si los datos de los que disponemos ya vienen unidos inseparablemente, como sucede en el caso de la cartografía impresa. La división en capas ofrece un mayor número de posibilidades distintas de trabajo y, como iremos viendo a lo largo de gran parte de este libro, también mayores posibilidades de análisis y proceso.

En resumen, el trabajo con capas permite una estructura más organizada y una mayor atomización de los datos, con las consecuentes ventajas en el almacenamiento, manejo y funcionalidad que esto conlleva.

4.6. Resumen

Los datos son una de las piezas más importantes del sistema SIG. Entendemos por dato un conjunto de valores o elementos que representan algo. La interpretación correcta de esos datos los dota de significado y produce *información*.

La información geográfica tiene dos componentes: una componente temática y una componente geográfica. Estas van unidas y conforman una unidad única de información geográfica, aunque pueden separarse y analizarse por separado. Mientras que la componente geográfica tiene un carácter fundamentalmente numérico, la componente temática puede incluir una o varias variables y estas ser de naturaleza muy variada.

La información geográfica se divide horizontal y verticalmente. Las unidades mediante las que incorporamos esta información a un SIG se conocen como *capas*, y son uno de los elementos primordiales en la estructura de manejo de datos de todo SIG. El trabajo con capas hace más transparente la gestión de la información geográfica en un SIG, permite una mejor integración de distintos datos, y es la base para muchas operaciones, algunas de las cuales iremos viendo en capítulos sucesivos.

Capítulo 5

Modelos para la información geográfica

5.1. Introducción

Los datos son, como ya sabemos, una parte imprescindible del SIG. Necesitamos conocer el área geográfica que estudiamos en un SIG (es decir, tener datos sobre ella), para así poder proceder a dicho estudio.

No obstante, convertir ese área geográfica y la información acerca de ella en un dato susceptible de ser incorporado a un SIG no resulta una tarea sencilla. Desde los orígenes de los SIG, una de las preocupaciones principales ha sido la de representar de la mejor manera posible toda la información que podemos extraer de una zona geográfica dada, de tal modo que pueda almacenarse y analizarse en el entorno de un SIG. Este proceso de representación, que ya desde el inicio planteaba problemas a los creadores de los primeros SIG, ha sido el responsable en gran medida de la arquitectura y forma de los SIG actuales, y a él se debe en buena parte el desarrollo que han experimentado tanto los SIG en sí como las disciplinas afines.

Describir los enfoques teóricos existentes para convertir la realidad relativa a una variable dada en una capa que la contenga de la forma más precisa posible y pueda ser empleada en un SIG es el objeto de este capítulo. Este proceso implica la construcción de un modelo (el dato geográfico), que representa la realidad y puede servir para conocer esta en profundidad a través de análisis que no se llevan a cabo sobre dicha realidad, sino sobre el modelo en sí.

El problema principal reside en el hecho de que el detalle real que encontramos en la naturaleza es prácticamente infinito, mientras que la representación y almacenamiento de esa realidad es finita. Se hace necesario extraer una serie de elementos y valores característicos, los cuales en ultima instancia se recogerán como valores numéricos dentro del SIG (pues son estos los que maneja un ordenador), y podrán interpretarse como el anteriormente citado modelo. El camino que lleva desde la realidad hasta ese conjunto de meros valores numéricos pasa por tres niveles:

- Establecimiento de un **modelo geográfico**. Es decir, un modelo conceptual de la realidad geográfica y su comportamiento.
- Establecimiento de un **modelo de representación**. Es decir, una forma de recoger el anterior modelo conceptual y sus características propias, reduciéndolo a una serie finita de elementos.
- Establecimiento de un **modelo de almacenamiento**. Es decir, un esquema de cómo almacenar los distintos elementos del modelo de representación.

El modelo geográfico es un ente puramente conceptual (de alto nivel), mientras que el de almacenamiento es más un concepto técnico inherente a la naturaleza informática del SIG (de bajo nivel)

5.2. Modelos geográficos

El primer paso hacia la creación del dato geográfico implica el establecimiento de un modelo conceptual relativo a cómo se ha de interpretar la realidad geográfica. Se trata de conceptualizar el espacio estudiado, la variable tratada y la variación de esta a lo largo del espacio. Este modelo geográfico es un esquema mental que constituye una forma particular de entender el hecho geográfico en sí, pero que todavía no incorpora elementos relativos a su representación o almacenamiento.

Existen muchos modelos geográficos distintos, entre los cuales cabe destacar dos de ellos [27]:

- Campos
- Entidades discretas

5.2.1. Campos

Un campo es un modelo de variación dentro de un marco n−dimensional, en el cual en cada punto dentro de dicho marco se tiene un valor de la variable estudiada. En su concepto matemático, un campo es una función de la forma $\varphi : \mathbf{R}^n \rightarrow \mathbf{R}^m$, esto es, una función que asocia cada punto de un espacio vectorial con otro en un espacio vectorial distinto.

En el caso más habitual, $m = 1$, es decir, que a cada punto del espacio vectorial origen se le asocia un único valor escalar. Se tiene así lo que se denomina un *campo escalar*. La mayoría de las variables que se emplean en un SIG necesitan un único valor para describirse (piénsese en variables como la elevación, la temperatura o la presión atmosférica, que solo requieren de un número para expresarse), por lo que los campos escalares son los más habituales en el ámbito geográfico.

No obstante, también encontramos los denominados *campos vectoriales*[1], en el cual el espacio vectorial de destino es multidimensional. Por ejemplo, para definir el movimiento del viento en un punto geográfico no basta con un único valor, sino dos: la velocidad y la dirección en la que sopla dicho viento. Dentro de un SIG, es habitual recoger los campos vectoriales como un conjunto de varios campos escalares, cada uno de ellos en una capa distinta. Así, se tendría una capa con la dirección y otra con la velocidad, ambas magnitudes escalares. Operando de esta manera, la solución no es única, ya que el vector resultante puede definirse mediante su módulo y dirección (como en el caso anterior), pero también por sus propias coordenadas en la base del espacio vectorial destino (en el caso anterior, las componentes x e y del vector que indica el movimiento del viento).

El espacio vectorial de origen puede ser bidimensional, es decir, una función de la forma $f(x, y)$, representando x e y las coordenadas geográficas. Este es el caso habitual en las capas que se emplean en un SIG, donde las variables que estudiamos adquieren uno u otro valor en función de su posición dentro de un sistema coordenado de referencia.

Puede añadirse una tercera dimensión, de tal modo que los valores dependan no solo de la posición sino igualmente de la elevación. Se tendría una función de la forma $f(x, y, z)$. Para el caso, por ejemplo, de la temperatura del aire, esta depende no solo de la localización, sino también de la altura. Otro ejemplo puede ser el porcentaje de

[1] El empleo del término *vectorial* para calificar a los campos vectoriales o los espacios vectoriales no debe confundirse con el modelo de representación vectorial que veremos más adelante en este capítulo. En el caso de campos y espacio, se trata de la terminología estándar del ámbito matemático, mientras que en el modelo de representación vectorial es una terminología propia de los Sistemas de Información Geográfica.

arena en el suelo, que depende de la localización pero también de la profundidad.

Igualmente, aunque en general es poco habitual en el marco de los SIG, puede añadirse la variable tiempo, teniéndose funciones de la forma $f(x, y, t)$ o $f(x, y, z, t)$

Por definición, un campo es continuo, ya que todos los puntos tienen un valor asociado. De igual modo, este valor es único, y no existe un elemento del espacio vectorial de partida que tenga asociados varios elementos del de destino, sean estos escalares o vectores.

Por su propia naturaleza los campos son ideales para modelizar variables que varían de forma continua en el espacio, entre ellas la practica totalidad de variables físicas del medio, tales como temperatura del aire, presión atmosférica, elevación, etc.

Los campos se asocian con las denominadas *coberturas*, termino este más empleado en el ámbito SIG. En una cobertura existe un valor único para todos los puntos de una región dada.

5.2.2. Entidades discretas

A diferencia de los campos, el modelo de entidades discretas no asocia a cada punto geográfico un valor, sino que concibe un entorno geográfico como un espacio vacío sobre el que se sitúan distintos elementos (entidades) que lo van rellenando. Cada una de dichas entidades posee unas características propias, constantes para toda ella, que son las que conferirán sus propiedades particulares a los puntos que se sitúen en su interior.

Un punto puede no pertenecer a ninguna entidad, o bien a varias de ellas, según sea la disposición de estas. Para un espacio dado, las entidades pueden ser todos aquellos elementos geométricos existentes en el mismo, tales como puntos, líneas, polígonos o, en el caso de ser dicho espacio de dimensión mayor que dos, también volúmenes.

Es fácil ver que el modelo de entidades discretas no es tan adecuado como los campos para conceptualizar variables continuas, ya que la continuidad de estas es opuesta al esquema discreto planteado. No obstante, otras variables no continuas se modelizan mejor mediante entidades discretas, ya que la forma en que se presentan coincide en cierta medida con dichas entidades como unidades mínimas.

La presencia de vías de comunicación, por ejemplo, se puede asimilar perfectamente a este modelo. Se tiene un espacio vacío (sin vías), en el cual se disponen los distintos viales en una serie de localizaciones concretas. Hay puntos que no estarán afectados por ninguna entidad, mientras que otros (los situados en las intersecciones) lo están por varias de ellas.

Las variables de tipo nominal y alfanumérico —las cuales no son, como vimos, continuas— tales como el tipo de suelo en un punto o el número de parcela catastral al que pertenece dicho punto, también se adaptan bien al modelo de entidades discretas.

Otra diferencia entre los campos y las entidades discretas es que estas últimas son en general más sencillas de comprender como concepto fuera de un ámbito técnico. Los campos son conceptos matemáticos que requieren un mayor grado de abstracción, y para la mayoría de la gente no resultan tan claros. Como algunos apuntan [28], el lenguaje habitual contiene un numero mayor de expresiones y recursos para describir la realidad geográfica en base a entidades discretas que en base a campos o conceptos abstractos similares.

5.3. Modelos de representación

Los modelos geográficos nos ofrecen una concepción particular del espacio geográfico y sus atributos. En base a ellos, el siguiente paso es

reducir las propiedades de dichos modelos a un conjunto finito de elementos, de tal modo que el registro de dichos elementos sirva para almacenar la realidad que los modelos geográficos describen. Para ello, empleamos los *modelos de representación*, también denominados *modelos de datos*.

Antes de entrar a describir los distintos modelos de representación, veamos algunos ejemplos que nos presentarán casos particulares de estos modelos, aclarando sus diferencias antes de proceder a una definición más detallada. En la figura 5.1 pueden verse distintas formas de representar la elevación de una zona, la cual, como ya sabemos, es una variable continua y puede concebirse mediante un campo escalar. Por el contrario, la red viaria se adapta mejor a un modelo de entidades discretas, y se muestran en la figura 5.2 sendas representaciones de esta variable según distintos modelos de datos. Mediante los ejemplos de estas figuras presentaremos los modelos de datos principales, así como su relación con los modelos conceptuales estudiados en el punto anterior.

Comenzando con la elevación, encontramos en la imagen cuatro distintas formas de representarla, a saber:

- **Curvas de nivel**. La representación clásica empleada tradicionalmente en los mapas de papel. Se recoge la elevación en una serie de curvas, que marcan los puntos en los que dicha elevación es múltiplo de una cierta cantidad (la equidistancia). En el ejemplo propuesto, se muestran curvas con elevaciones múltiplos de 10 metros.
- **Una malla de celdas regulares**, en cada una de las cuales se dispone un valor, que corresponde a las características de la zona ocupada por dicha celda. En este caso, cada celda tiene un valor de altura propio, que al convertirse en un color mediante el uso de una escala de colores, da lugar a la imagen mostrada.
- **Puntos regulares**. Una serie de puntos regularmente espaciados. Existe información de la elevación solo en dichos puntos. La información se muestra como etiqueta asociada a cada punto.
- **Red de Triángulos Irregulares**. Una Red de Triángulos Irregulares (TIN en sus siglas inglesas, de *Triangulated Irregular Network*), es una estructura en la cual se toman los puntos más característicos del relieve y en base a ellos se construye una teselación en triángulos con unas condiciones particulares. Cada uno de los triángulos posee unas propiedades comunes en cuanto a su relieve. Veremos más adelante en detalle este tipo de estructuras. Por el momento, basta recordar que los elementos básicos de esta forma de representación son triángulos.

Para el caso de las vías encontramos dos representaciones distintas:

- **Una malla de celdas** como la citada en el caso anterior. Las celdas de vía tienen un valor (representado aquí en azul) distinto de las que se encuentran fuera de la vía (representado en blanco)
- **Un conjunto de líneas** representando los trazados de las vías.

En este ultimo caso las celdas se han elegido de un tamaño excesivamente grande, con el fin de que pueda apreciarse de forma inmediata la diferencia existente. Veremos más adelante que, como no es difícil intuir, la representación mediante celdas no es tan adecuada para el caso de una capa de vías (aunque para el caso de la elevación da lugar a una imagen con un aspecto inmejorable y altamente informativo), cuando estudiemos los aspectos relativos a la precisión en los distintos modelos de almacenamiento.

Como vemos, para un mismo tipo de información existen diversas alternativas en cuanto a la forma de materializar la realidad y plasmar el modelo geográfico concreto. Estas formas las podemos clasificar en dos grupos principales: mo-

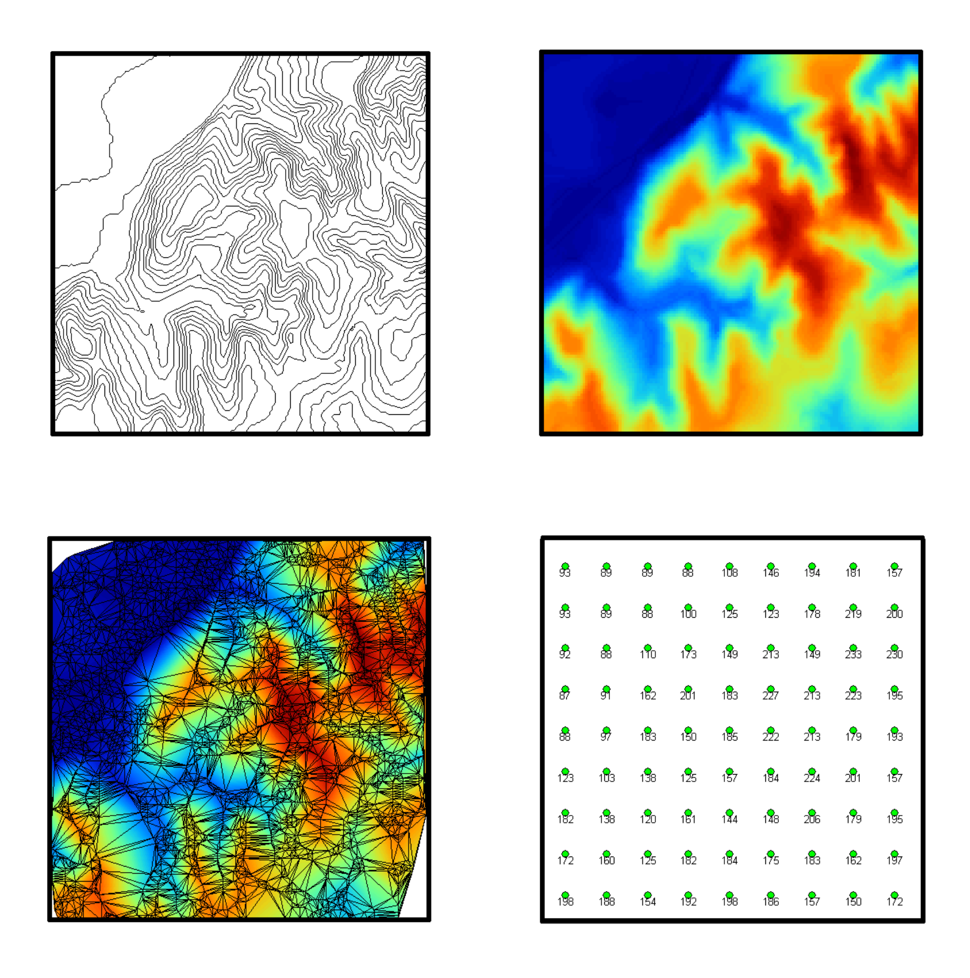


Figura 5.1: Distintas formas de representar una capa con información altitudinal.

delo de representación *ráster* y modelo de representación *vectorial*.

Si se han seguido los capítulos de partes anteriores, probablemente los términos *ráster* y *vectorial* no resulten extraños, ya que han aparecido con cierta frecuencia. Esto es así porque, además de definir dichos términos los principales modelos de representación de la información geográfica dentro de un SIG, se han venido utilizando tradicionalmente para definir a los SIG en sí, en función de si sus capacidades se hallaban más enfocadas al manejo y análisis de información en formato ráster o en formato vectorial. A día de hoy, esa diferencia no es tan patente y los SIG más habituales pueden trabajar con ambos indistintamente, pudiendo realizar las tareas que resultan más adecuadas de llevar a cabo tanto con uno como con otro tipo de representación.

En líneas generales podemos decir que el modelo ráster se basa en una división sistemática del espacio, la cual cubre todo este (a este concepto se le denomina *teselación*), caracterizándolo como un conjunto de unidades elementales (las celdas de las mallas vistas en los ejemplos). El modelo vectorial, por su parte, no divide el espacio completamente, sino que lo define mediante

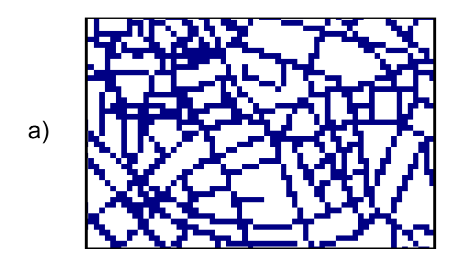




Figura 5.2: Distintas formas de representar una capa con información sobre una red viaria.

una serie de elementos geométricos con valores asociados, siendo la disposición de estos no sistemática, sino guardando relación con los objetos geográficos presentes en la zona de estudio.

En un principio, puede pensarse que el modelo ráster se asemeja al modelo geográfico de campos, mientras que el vectorial concuerda con el de entidades discretas. Aunque en cierta medida puede considerarse que así sucede y existe tal dualidad, no es del todo cierta esta equiparación, como discutiremos con algo más de detalle en los siguientes puntos.

De forma esquemática, los enfoques de los modelos de representación ráster y vectorial se muestran en la figura 5.3

Podemos entender estos enfoques haciendo uso del esquema de Sinton presentado con anterioridad. En el modelo vectorial controlamos la definición de los valores asociados, y medimos la localización y forma de estos, dejando fijo el tiempo. En el modelo ráster, aunque la componente temporal también es fija, la componente que controlamos es la espacial (a través de la sistematicidad de la malla), mientras que medimos la naturaleza de los valores en cada una de las celdas.

Antes de pasar a la definición detallada de los modelos ráster y vectorial, mencionar que, como modelos principales empleados para la definición de capas de información geográfica, las expresiones *capa vectorial* y *capa ráster* son de uso habitual, y se emplearán de aquí en adelante tanto en este como en posteriores capítulos.

5.3.1. Modelo ráster

En el modelo ráster, la zona de estudio se divide de forma sistemática en una serie de unidades mínimas (denominadas habitualmente *celdas*), y para cada una de estas se recoge la información pertinente que la describe. Se puede ver esto en detalle en la figura 5.4, que muestra aumentada una porción la malla ráster de elevaciones de la figura 5.1, de modo que los límites de las celdas se hacen patentes y puede además representarse en cada una de ellas su valor asociado.

Aunque la malla de celdas puede contener información sobre varias variables, lo habitual es que trate una única variable. Es decir, que se tenga un único valor para cada una de las celdas.

La característica principal del modelo ráster, y que le confiere gran parte de sus propiedades más interesantes, especialmente de cara al análisis, es su sistematicidad. La división del espacio en unidades mínimas se lleva a cabo de forma sistemática de acuerdo con algún patrón, de tal modo que existe una relación implícita entre las celdas, ya que estas son contiguas entre sí, cubren todo el espacio, y no se solapan. Por tanto, la posición de una celda depende de la de las restantes, para así conformar en conjunto toda la malla regular que cumple las anteriores características. Dicho de otro modo, el orden propio de

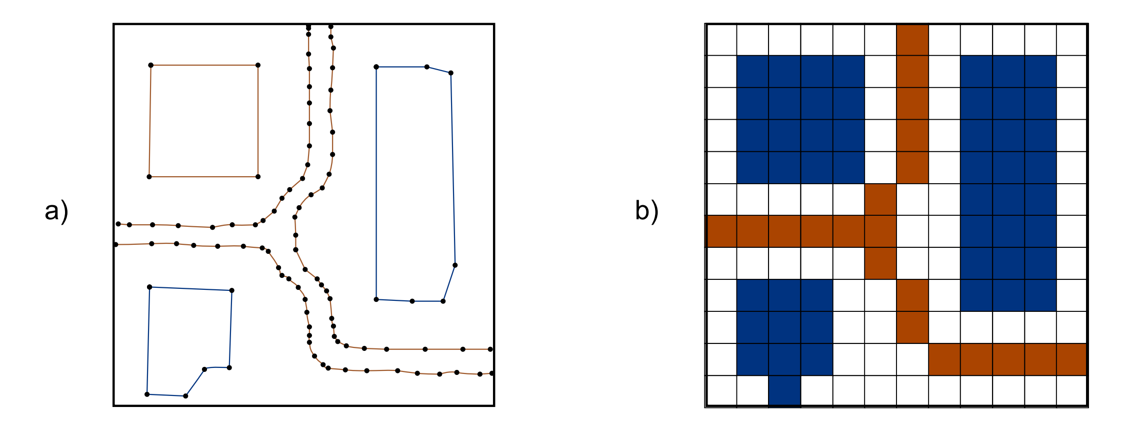


Figura 5.3: Comparación entre los esquemas del modelo de representación vectorial (a) y ráster (b).

132.90	133.90	135.60	139.20	140.60	144.40
126.80	128.50	131.40	133.90	135.90	139.90
122.30	124.10	125.80	128.50	131.80	135.60
118.70	119.90	121.90	124.30	128.20	131.50

Figura 5.4: Celdas de una malla ráster con sus valores asociados.

las celdas, presente gracias a la división sistemática realizada, aporta un elemento adicional que las relaciona entre sí.

Como unidad mínima pueden tomarse elementos de diversas formas. La más habitual es mediante unidades de forma cuadrada, aunque también pueden ser formas rectangulares, o incluso triangulares o hexagonales [29]. No obstante, los SIG habituales se limitan a modelos de celdas cuadradas o rectangulares, y las implementaciones de otros modelos son de uso muy reducido y en aplicaciones muy especificas que en general no están orientadas al uso general. Junto a esto, la información geográfica en formatos ráster distintos de la división en celdas cuadradas o rectangulares es prácticamente inexistente, haciendo más difícil el empleo de estos formatos en condiciones normales de trabajo.

De igual modo, existen representaciones ráster no regulares, en las que todas las unidades mínimas no tienen un mismo tamaño. Este tipo de representaciones no tiene apenas presencia en los SIG, pero son habituales en otros ámbitos tales como el de las representaciones 3D, con unos requerimientos bien distintos[2]. Esto está relacionado a su vez con los modelos de almacenamiento ráster, que veremos más adelante en este mismo capítulo.

[2]Véase, por ejemplo, el concepto de Nivel Continuo de Detalle (Continuous Level of Detail, CLOD), para lograr representaciones de detalle con el menor gasto de recursos posible, y que es habitual en este campo.

En todos los casos, la división en celdas no depende de la variable estudiada, y es una división geográfica. Esto lo diferencia de otras divisiones como el caso de la Red de Triángulos Irregulares, que, a pesar de ser una teselacion que cubre todo el espacio, está basada en la propia variable de elevación, y dicha división (número, forma y disposición de los triángulos) sería distinta en caso de que los valores de elevación fueran otros.

Siendo las mallas ráster de celdas cuadradas las más habituales, pasemos a ver algo más acerca de ellas. Dos son los elementos principales que resultan necesarios para una definición completa de una capa ráster:

- Una localización geográfica exacta de alguna celda y una distancia entre celdas, para en base a ellas, y en virtud de la regularidad de la malla, conocer las coordenadas de las restantes.
- Un conjunto de valores correspondientes a las celdas.

En el modelo ráster no se recogen de forma explícita las coordenadas de cada una de las celdas, sino tan solo sus valores. No resulta necesario acompañar a dichos valores de un emplazamiento espacial concreto, pues hacen referencia a un elemento particular de la malla, la cual representa una estructura fija y regular. No obstante, sí que es necesario emplazar dicha malla en el espacio para después poder calcular las coordenadas particulares de cada celda.

Lo más habitual es definir el emplazamiento de una única celda (habitualmente la celda superior izquierda), una orientación fija, y una distancia entre las celdas (el paso de la malla). Como se muestra en la figura 5.5, esto ya permite, mediante un sencillo cálculo, conocer las coordenadas de todas las celdas sin necesidad de almacenar estas.

La orientación de las capas ráster es habitualmente Norte–Sur, de tal modo que si pasamos de la primera a la segunda fila estamos descendiendo en latitud. Dicho de otra forma, la parte de arriba de la imagen es el Norte, y la de abajo es el Sur. Esta convención simplifica el trabajo con capas ráster dentro de un SIG y permite aplicar directamente la fórmula mostrada en la figura 5.5.

Puede suceder, aun así, que la fuente de datos original no se adhiera a este formato (por ejemplo, una fotografía aérea en la que el avión no volaba en dirección Norte–Sur o perpendicular, o una porción de un mapa escaneado que no tiene tampoco esa orientación). En tal caso, y puesto que los SIG trabajan en general con esa orientación en sus representaciones y a la hora de incorporar capas ráster, nos encontraremos con situaciones como la mostrada en la figura 5.6

En ella vemos cómo la orientación de la banda de estudio recogida es distinta de la Norte–Sur de la imagen. Esto, unido a la forma rectangular que ha de tener dicha imagen, causa la aparición de zonas sin información (en negro). Esto implica por una parte la necesidad de almacenar un gran número de valores sin interés, y por otra la necesidad de especificar de algún modo que todas esas celdas que aparecen en negro en la imagen son realmente celdas para las cuales no se dispone de información. Esto último se suele llevar a cabo mediante la definición de un valor arbitrario que indique la falta de datos (denominado generalmente valor de *sin datos*), que codifica tal situación, de tal modo que pueden ignorarse las celdas con dicho valor a la hora de representar o analizar la capa ráster en cuestión.

El otro parámetro necesario junto con la orientación de la malla y la situación geográfica de una de sus celdas es el denominado *tamaño de celda* o *tamaño de píxel*, también conocido como *resolución*, pues, en efecto, su magnitud define la resolución de la capa. Un tamaño de celda mayor implica una menor resolución, y viceversa.

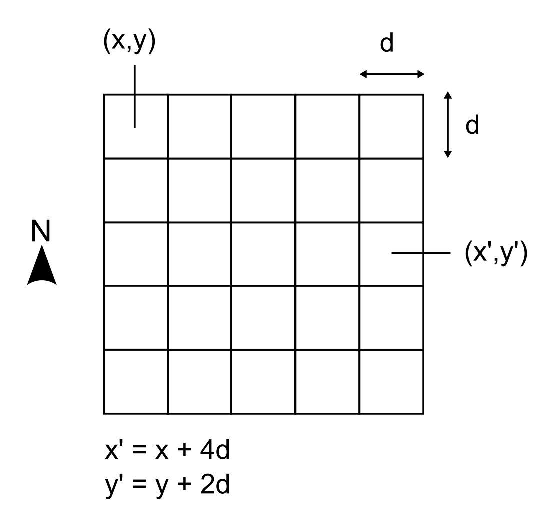


Figura 5.5: La estructura regular de la malla ráster permite conocer las coordenadas de las celdas sin necesidad de almacenar estas, sino tan solo recogiendo algunos parámetros de la malla como la localización de una celda base (x, y), la orientación global o el tamaño de celda (d).

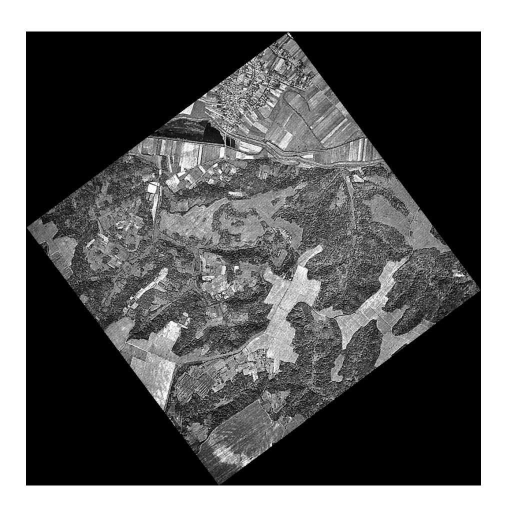


Figura 5.6: Aunque la zona de estudio no tenga orientación Norte–Sur, los SIG trabajan habitualmente con esta orientación, y las imágenes deben adecuarse a ello.

Además de servir para el cálculo de coordenadas de las celdas y definir la estructura de la malla, el tamaño de celda permite calcular áreas, ya que establece el área ocupada por cada celda. Asimismo, y como aspecto más relevante, el tamaño de celda determina la precisión con la que se recoge una variable dentro de una capa ráster, y puede considerarse como el equivalente conceptual a la escala de dicha capa. Por esta razón, es importante trabajar con capas ráster de un tamaño de celda adecuado para el tipo de análisis o tarea que quiera desarrollarse.

Así, un análisis microtopográfico en el cual resulta necesario registrar la variación del relieve con detalle no puede llevarse a cabo con una capa de elevaciones con tamaño de celda de 100 metros, ya que toda la variabilidad menor a esos 100 metros se pierde. No debe olvidarse que cada celda registra un único valor de la variable, y esta se considera constante dentro de dicha celda. Un tamaño de 100 metros implicaría tomar un único valor para cada hectárea de terreno, lo cual no es suficiente en este caso.

Los factores que influyen en el tamaño de celda de una capa ráster son muchos, entre ellos las características de los datos iniciales con los que se ha creado dicha capa o los medios particulares con que estos se han recogido. En la figura 5.7 pueden observarse dos imágenes aéreas (las imágenes son un tipo particular de capa ráster, como en breve veremos) con distinta resolución. Esta, al ser distinta, las hace válidas para uno u otro tipo de uso. Vemos claramente que en en la imagen en blanco y negro (cuyo tamaño de píxel es de 5 metros) se distinguen las distintas áreas de cultivo, mientras que en la imagen en color (con tamaño de píxel de 25 metros), estos no se distinguen. Todos aquellos análisis que requieran

disponer de información por debajo de esos 25 metros, no podrán ser llevados a cabo con esta última imagen.

Para el caso de capas ráster de variables continuas, en la sección 14.2.8 se da información detallada sobre cómo definir el tamaño de celda óptimo a la hora de crear estas a partir de datos de otra clase tales como datos vectoriales.

Una vez conocemos el formato ráster, podemos relacionarlo con lo que ya hemos visto relativo a los modelos geográficos. En primer lugar, y por sus propias características, puede pensarse que la representación ráster es más adecuada para variables de tipo continuo que varían a su vez de forma continua en el espacio geográfico. Es decir, es más próxima al modelo geográfico de campos que al de entidades discretas. Esto es así debido principalmente a que una capa ráster cubre todo el espacio, y ello favorece el estudio de dicha variabilidad. A pesar de ello, no debe considerarse que el ámbito de las variables continuas y los campos es exclusivo de las capas ráster. De hecho, de las cuatro representaciones mostradas para el caso de la elevación, solo una de ellas es de tipo ráster.

Sí es cierto, no obstante, que el formato ráster es especialmente adecuado para el análisis de la información geográfica, en especial cuando esta es de tipo continuo. Esto es así porque el principal elemento de las capas ráster es, como ya se ha dicho, su estructura sistemática. Si a esta le unimos la regularidad que se presenta en la forma más extendida de representación ráster (la de celdas cuadradas), tenemos un modelo óptimo para el análisis, que simplifica en gran medida este y hace más sencilla la implementación de los algoritmos correspondientes. Es por ello que, tradicionalmente, los SIG con mayor soporte para datos ráster han sido aquellos que presentaban a su vez un mayor número de funcionalidades de análisis en áreas tales como el estudio del relieve, el análisis de costes u otros similares.

El caso de las imágenes

Un caso especial de capa ráster son las imágenes, de las que hemos visto ya un ejemplo al tratar el tamaño de celda. Tanto si estas proceden de un sensor digital o bien han sido escaneadas, los sensores correspondientes generan una estructura en forma de malla que se ajusta al modelo de representación ráster. Este hecho tiene gran importancia, pues facilita el análisis conjunto de imágenes y capas de datos con otro tipo de información, haciendo que este sea sumamente más sencillo, al compartir el modelo de representación.

Mientras que, como hemos visto en los ejemplos, una misma información se puede recoger en formatos ráster y vectorial, las imágenes se recogen únicamente en formato ráster, tanto por ser ese modelo mucho más adecuado, como por ser mucho más coherente con el tipo de información y la procedencia de esta.

El concepto de celda en una malla ráster es el equivalente al de píxel[3], bien conocido en el campo de las imágenes digitales. Así, cuando decimos que una cámara digital tiene tres megapíxeles, queremos decir que captura un total de tres millones de píxeles. De otra forma, la malla ráster que se genera tiene tres millones de celdas. Las imágenes con las que trabajamos en un SIG no se diferencian de las que tomamos con una cámara digital, salvo en el hecho particular de que representan una porción de terreno dentro de un sistema de coordenadas dado, pero la estructura es la misma: una malla de celdas (píxeles).

Otra particularidad de las imágenes es la presencia de *bandas*. Los valores recogidos en las imágenes indican de forma general la reflectan-

[3]acrónimo de *picture element*





Figura 5.7: Imágenes de diferente resolución en función del sensor con que han sido obtenidas. Al tener distintos tamaños de píxel, servirán para distintos usos dentro de un SIG.

cia en una determinada longitud de onda (esto se explica con mayor detalle en los capítulos 6 y 17). Puesto que el espectro de radiación puede subdividirse en distintos grupos, los sensores que toman estas imágenes recogen varias capas, una para cada uno de estos grupos. En lugar de almacenarse como un conjunto de capas separadas, es más frecuente que lo hagan en una única que contiene varias *bandas*, es decir, varios niveles distintos, cada uno de los cuales podría constituir por sí mismo una capa ráster.

Aunque se trata de una diferencia más de tipo formal, tiene cierta importancia, ya que no todos los SIG están preparados para manejar capas ráster con independencia de su número de bandas. Imágenes con una única banda, o tres, son habituales y soportadas en la mayoría de implementaciones, pero muchos programas no están preparados para soportar capas con un mayor número de bandas.

Todos estos conceptos se detallarán en el capítulo 6.

5.3.2. Modelo vectorial

El otro modelo principal de representación es el modelo vectorial. En este modelo, no existen unidades fundamentales que dividen la zona recogida, sino que se recoge la variabilidad y características de esta mediante entidades geométricas, para cada una de las cuales dichas características son constantes. La forma de estas entidades (su frontera), se codifica de modo explicito, a diferencia del modelo ráster, donde venía implícita en la propia estructura de la malla.

Si el modelo ráster era similar al modelo conceptual de campos, el vectorial lo es al de entidades discretas, pues modeliza el espacio geográfico mediante una serie de primitivas geométricas que contienen los elementos más destacados de dicho espacio. Estas primitivas son de tres tipos: puntos, líneas y polígonos.

Utilizando puntos, líneas o polígonos, puede modelizarse el espacio geográfico si se asocia a estas geometrías una serie de valores definitorios. La componente espacial de la información queda así en la propia primitiva (recoge la forma, posición y otras propiedades espaciales), y

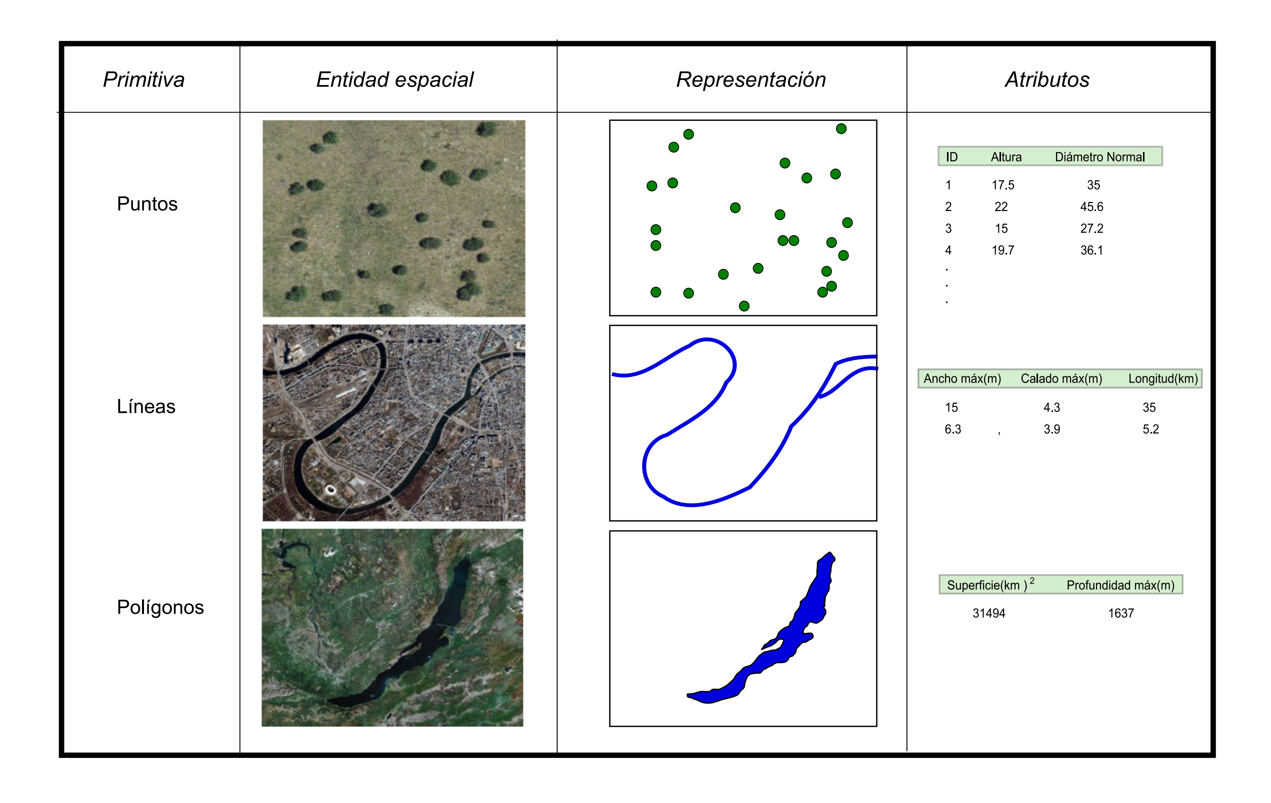


Figura 5.8: Primitivas geométricas en el modelo de representación vectorial y ejemplos particulares de cada una de ellas con atributos asociados

la componente temática queda en dichos valores asociados (Figura 5.8).

A la hora de definir las formas geométricas básicas, todas ellas pueden reducirse en última instancia a puntos. Así, las líneas son un conjunto de puntos interconectados en un determinado orden, y los polígonos son líneas cerradas, también expresables por tanto como una serie de puntos. Todo elemento del espacio geográfico queda definido, pues, por una serie de puntos que determinan sus propiedades espaciales y una serie de valores asociados.

Una única entidad (para la cual existirá un único conjunto de valores asociados) puede contener varias primitivas. Así, en un mapa mundial en que cada entidad represente un país, y tal y como se ve en la figura 5.9, países como Canadá estarán representados por más de un polígono, pues no puede recogerse todo su territorio mediante uno único. Todos estos polígonos constituyen una única entidad, ya que todos pertenecen al mismo país y tendrán el mismo conjunto de valores asociados.

Otro caso particular en las capas de polígonos son aquellos polígonos con islas (huecos). En este caso, se registran de la misma forma que en el caso de varios polígonos disjuntos. Se recogen los propios huecos como polígonos independientes, pero recogiendo de algún modo también la circunstancia de que estos polígonos no se *suman* a los polígonos existentes en esa entidad, sino que se *restan*. Así es, por ejemplo, para el caso

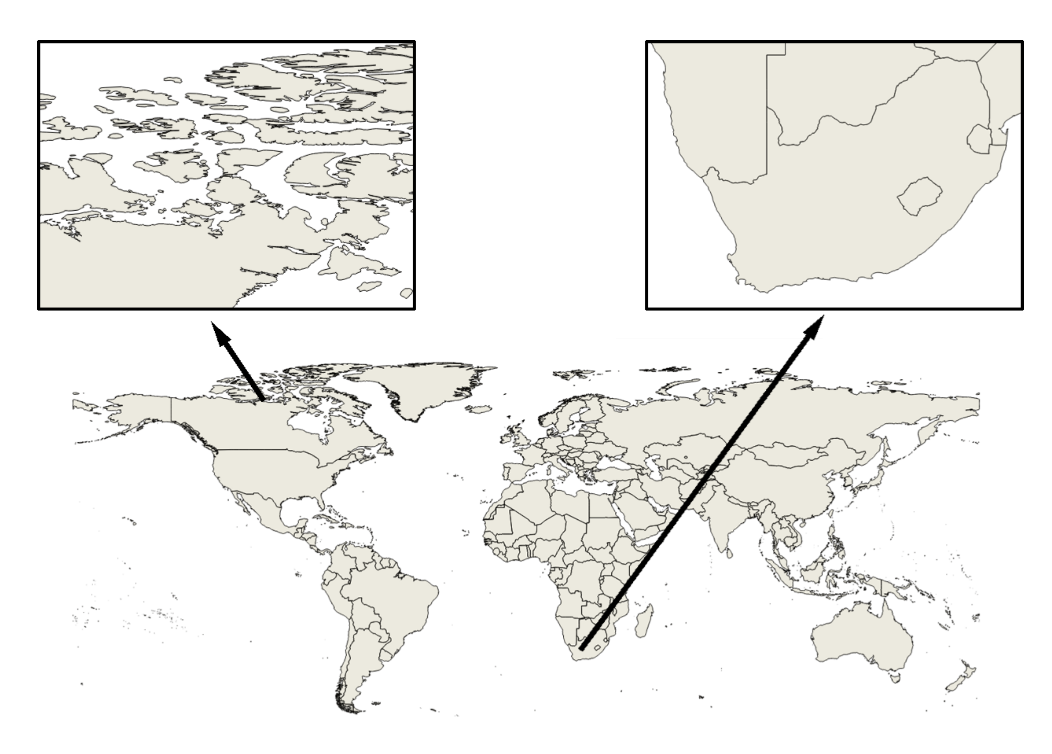


Figura 5.9: Casos particulares de polígonos: a) varios polígonos disjuntos en una misma entidad (en este caso, mismo país), b) Polígonos con islas (huecos).

del área total de polígonos de una única entidad, ya que el área del hueco debe ser restada de la total.

En la figura anterior, vemos como Sudáfrica presenta esta situación, ya que dentro del territorio del país hay zonas aislada que no pertenece a Sudáfrica, como por ejemplo la que constituye el Reino de Lesotho.

Como se muestra en la figura 5.10, el conjunto del territorio ocupado por Sudáfrica y las zonas interiores que no pertenecen al país no puede verse como un conjunto de polígonos sin más. Para representar Sudáfrica de forma aislada es necesario «restar» del polígono que engloba todo el territorio los polígonos respectivos a los países interiores. De no hacerlo así, un cálculo sencillo tal y como el del área de dicho país arrojará un resultado erróneo, pues considerará igualmente estas zonas interiores.

En realidad, los huecos se registran como polígonos disjuntos que pertenecen a la entidad, aunque en lugar de representar un territorio que se añade, representan uno que se «quita». Una forma habitual de hacer esto es almacenar las coordenadas de los vértices de estos polígonos interiores en sentido inverso, de tal modo que su área es negativa. De esta forma, la suma total del área de los polígonos de la entidad es igual al área buscada[4].

Dentro de un SIG, una capa vectorial puede contener un único tipo de primitiva. Así, tenemos capas vectoriales de puntos, de líneas y de polígonos, respectivamente. La elección de uno u otro tipo de capa para registrar una variable o

[4]La fórmula empleada para el cálculo del área de un polígono se expone en la página 212

conjunto de ellas ha de ser función del tipo de fenómeno que se pretende modelizar con dicha capa o la precisión necesaria, entre otros factores.

Por ejemplo, una capa de puntos puede representar un conjunto de ciudades, cada una de ellas definida como un único punto. Sin embargo, puede emplearse una capa de polígonos y no recoger una única coordenada (correspondiente, por ejemplo, al centro de la ciudad), sino el contorno o los límites administrativos de esta. Dependiendo del caso, será más apropiado elegir una u otra alternativa.

De igual modo, la capa de vías representada en la figura 5.2 es una capa de líneas. Cada línea, como elemento teórico de ancho nulo, representa el eje de la vía. Si se requiere una mayor precisión en la definición de la superficie de rodadura de dichas vías, una capa de polígonos puede ser utilizada en lugar de una de líneas.

Lo anterior tiene una evidente relación con los conceptos de escala y generalización que vimos en el capítulo 3.

No debe pensarse que las capas vectoriales, sean del tipo que sean, se emplean únicamente para recoger fenómenos o elementos cuya forma coincide con la de las primitivas geométricas (es decir, puntos para recoger elementos puntuales, líneas para aquellos elementos con una dimensión mucho menor que la otra, y polígonos para el caso de superficies). Además de los ejemplos anteriores, debemos recordar que el modelo vectorial también sirve para representar campos y recoger variables tales como la elevación.

Así, en los ejemplos de la figura 5.1 encontramos capas de puntos, líneas (curvas de nivel) y polígonos (TIN), todas ellas empleadas para representar la variable elevación. En ocasiones se emplean las primitivas para recoger objetos reales de forma similar, mientras que en otros casos sirven para plantear un modelo lógico y recoger variables que no se asemejan de modo alguno a las formas geométricas registradas.

A propósito de la capa de puntos regulares, cabe pensar que es similar en concepto y forma a la malla ráster, ya que es regular. Sin embargo, existen dos diferencias importantes: en primer lugar, en la capa de puntos hay zonas en blanco, de las que no sabemos su elevación, mientras que en la malla ráster las celdas tienen una superficie y cubren en su conjunto todo el espacio. En segundo lugar, si tenemos esa capa de puntos en un SIG, esta va a contener las coordenadas particulares de cada punto, ya que en sí las capas vectoriales no son regulares (pueden guardar alguna regularidad, pero no necesariamente), y por tanto es necesario, como hemos visto, registrar explícitamente sus coordenadas. De modo similar podríamos hacer una capa de polígonos cuadrados, pero seguiría sin ser una malla ráster, más aún si careciera de un elemento que veremos en breve: la topología.

La componente temática en el modelo vectorial

Hemos visto que la forma en la que los modelos de representación separan las dos componentes de la información geográfica es bien distinta. En el modelo ráster se tiene un conjunto de valores (la componente temática), los cuales guardan una estructura dada, la cual por sí misma establece su disposición en el espacio (la componente espacial). En el vectorial, por su parte, la componente espacial se recoge explícitamente según una serie de puntos, la cual puede ser más o menos compleja en función de la complejidad de la entidad a representar o el detalle con que se recoja. A este conjunto de puntos se le relaciona después con una serie de valores, que son los que definen las propiedades de la entidad.

Estos valores, los *atributos*, a diferencia del caso ráster, suelen ser múltiples. Por ejemplo,

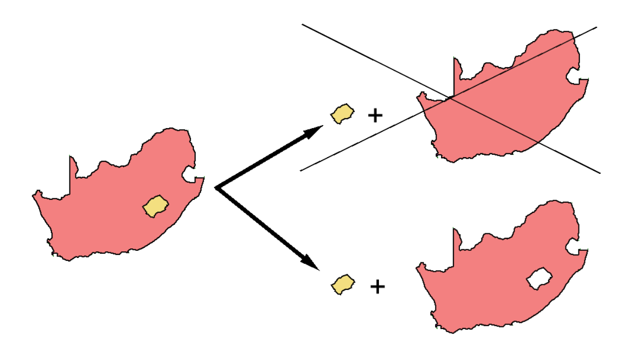


Figura 5.10: Los huecos de un polígono han de considerarse como parte de este.

dada una capa vectorial de países, podemos recoger valores asociados a cada país tales como su superficie, su población, el Producto Interior Bruto, el nombre de su capital o su idioma oficial. Todo este conjunto de valores se asocian a una única copia de la componente espacial, y esta no debe repetirse para recoger cada uno de esos parámetros. En el modelo ráster, si tenemos n capas distintas, en realidad estamos almacenando n veces la componente espacial.

Por esta estructura particular, la componente temática se presta especialmente a almacenarse en una base de datos, siendo en la actualidad las más extendidas las denominadas *bases de datos relacionales*. Estas bases de datos se *enlazan* a la componente espacial y permiten una serie de operaciones(ver capítulo 12) y un manejo ventajoso de los *atributos*. Existen, por tanto, dos realidades: la relativa a la componente geográfica y la base de datos que gestiona los atributos, la cual permite análisis y operaciones independientes, del mismo modo que si no existirá una localización asociada a dichos atributos. Estas realidades pueden estar muy separadas, gestionándose en aplicaciones distintas y almacenándose en ficheros diferentes, con lo cual existe una división formal mucho más acusada que en el caso de las capas ráster, que se asemejan más a unidades de información autocontenidas.

En el caso de las capas ráster, no es necesario recurrir a una base de datos, y simplemente la representación del conjunto de valores de la variable en las distintas celdas sirve para el almacenamiento, análisis y manejo de la información. Como indica [11], esta forma de *conectar* las componentes espacial y temática es apta para el análisis, pero el manejo de los atributos requiere la presencia de una base de datos.

El establecimiento de las bases de datos, su manejo y su implementación dentro de un SIG es un tema altamente complejo. La forma en que el manejo de la componente temática y la gestión de la base de datos se establecen, así como la imbricación de la una en la otra, es la materia exclusiva del capítulo 8, donde todos estos temas se desarrollarán con profundidad.

Topología

Un elemento particular del modelo de representación vectorial es la *topología*. En términos matemáticos, la topología estudia las características de los objetos geométricos que no varían al aplicar una transformación topológica tal como, por ejemplo, una transformación afín. Si tomamos un mapa y lo distorsionamos, los ángulos,

las superficies y las distancias se ven afectadas. Sin embargo, otras propiedades tales como la adyacencia entre elementos o las relaciones entre estos se conservan. Por ejemplo, si una ciudad está dentro de una determinada provincia en un determinado mapa, no existe forma de distorsionar esta para lograr que dicha ciudad se encuentre fuera de la provincia.

En el ámbito de los SIG se entiende la topología desde un punto de vista menos estricto y más funcional. En general, se dice que una capa tiene topología si en ella se almacenan de algún modo las relaciones entre los distintos elementos que la componen. En caso contrario, la capa es de tipo puramente cartográfico, ya que los elementos que contiene no presentan relación entre sí, o al menos esta relación no está almacenada junto a la propia información de estos elementos.

En una capa ráster, las relaciones topológicas vienen implícitas en el propio modelo ráster, y son ajenas a la información como tal, dependiendo de la estructura de la malla de datos en sí. En el modelo vectorial, sin embargo, se recogen valores relativos a cada elemento de forma individual, y si las relaciones existentes no se registran de modo explicito, no se tendrá posteriormente información sobre ellas.

Disponer de topología en una capa vectorial es de gran importancia a la hora de llevar a cabo ciertos tipos de análisis, así como otras operaciones tales como la edición de los propios datos geográficos. La topología no aporta beneficio a la hora de representar una capa, pero sí a la hora de llevar a cabo un análisis sobre ella [30].

En la figura 5.11 se puede observar la diferencia existente entre editar una capa de polígonos con topología y una sin ella. En el primer caso, la información contenida en la capa antes de su edición nos informa no solo de la forma de cada polígono, sino también del hecho de que ciertos polígonos comparten bordes comunes y de que el conjunto de ellos cubre el espacio de forma completa (constituyen una teselación). Así, al modificar un punto en uno de los polígonos, todos aquellos polígonos adyacentes que comparten dicho punto modifican también su perímetro. Las capacidades de edición implementadas en el Sistema de Información Geográfica hacen uso de la información topológica a la hora de editar geometrías. En el segundo caso, sin embargo, esta información no existe, y al modificar un punto el SIG no sabe cómo alterar los polígonos adyacentes, perdiéndose la teselación completa del espacio.

La topología es en este caso un elemento que contribuye a la calidad de los datos, pues mantiene la coherencia espacial de estos y evita la aparición de elementos tales como polígonos de muy pequeño tamaño, frecuentes en la digitalización de entidades debido a las pequeñas imprecisiones que se presentan en el proceso, y que causan la presencia de falsos solapes entre polígonos.

No obstante, no todos los SIG incorporan capacidades de manejo y análisis de capas vectoriales con topología, y son menos aún los que implementan capacidades para crear dicha topología. En general, estas han quedado reservadas a las aplicaciones de alta gama, y el manejo de información vectorial en los SIG de escritorio no incluye de forma general lo relativo a la topología.

Otro ejemplo de proceso en el que se hace necesario el disponer de capas con topología es el análisis de redes (este se detalla en el capítulo 20.7). Un mero conjunto de elementos geométricos (líneas en este caso) no nos da información sobre los posibles enlaces entre las vías que quedan representadas. Los puntos donde se cruzan dos vías pueden ser cruces o rotondas (es decir, puede pasarse de una vía a otra, existiendo conexión entre ellas), o bien pasos elevados o subterráneos donde una de las vías pasa por encima de

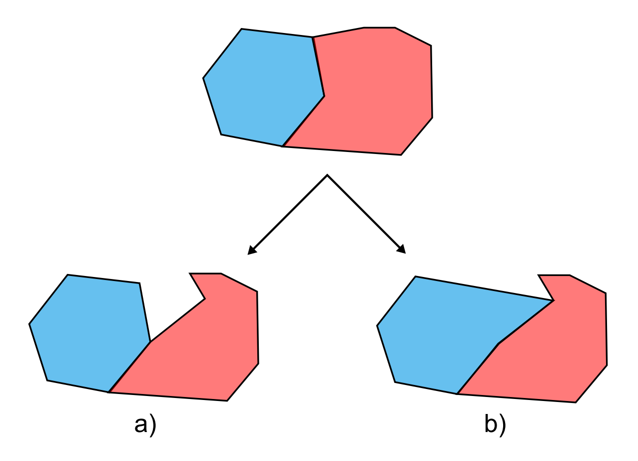


Figura 5.11: Diferencias entre la edición (desplazamiento de un punto) no disponiendo de topología (a) o con ella (b).

la otra (y por tanto no existe comunicación entre ambas). Las circunstancias son muy distintas en función del tipo de cruce que exista, y por ello es imprescindible conocer esta información para efectuar un análisis de redes correcto.

Otro elemento que no se puede recoger sin topología son las direcciones de circulación. Habrá vías que puedan recorrerse en ambos sentidos, mientras que habrá otras que solo permitan movimiento de tráfico en una dirección. Saber en qué dirección podemos recorrer una vía es vital para poder plantear cualquier tipo de análisis, y esta es una información de la que no disponemos si nuestra red viaria no ha sido representada mediante un modelo con topología.

La figura 5.12 recoge de forma esquemática estas circunstancias.

Aunque, como se ha mencionado, las capas ráster en cierta forma contienen información topológica (se conoce la relación de adyacencia entre las distintas celdas), esta es *débil* y no suficiente para análisis complejos como el de redes donde existen distintos elementos como los mencionados cruces o las direcciones de circulación. Aparte de la inherente peor disposición del modelo de representación para recoger una entidad espacial tal como una red, el modelo ráster no es óptimo para recoger la necesaria información topológica al respecto. Existen algunos intentos de adaptarlo a estas circunstancias (véase, por ejemplo [31]), pero en general no se encuentran implementados de forma habitual.

Modelo vectorial sin topología (*spaguetti*)

En el tipo más simple de modelo vectorial, únicamente se recogen las propiedades geométricas de cada entidad, almacenando para cada una de ellas el conjunto de puntos individuales que la componen. Esto aporta toda la información necesaria sobre la entidad, pero deja de lado la topología. Algunas propiedades topológicas pueden calcularse, tales como saber si un punto esta contenido dentro de un polígono o si dos rectas se cruzan, pero para otras no se dispone de información suficiente. Así, aunque podamos saber si dos líneas se cruzan, no podemos saber si este cruce implica una conexión real entre ellas de forma que pueda pasarse de la una a la otra o no, como vimos en la figura 5.12.

Esta forma de recoger las entidades vectoriales es similar a la que encontramos en un mapa

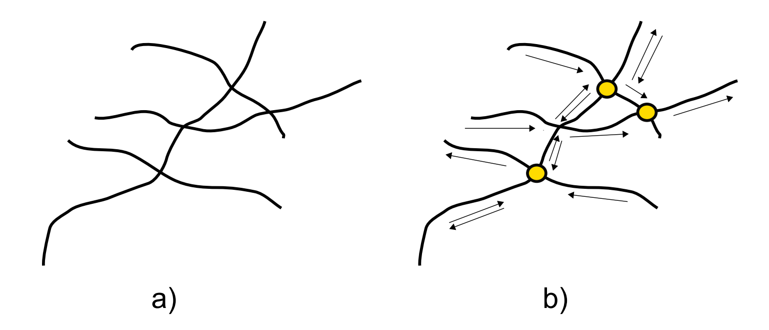


Figura 5.12: Capa de vías de comunicación sin topología (a) o con ella (b). Los puntos en este segundo caso indican conexiones entre vias, y son una representación visible de la topología existente. Las flechas indican la dirección de circulación y, al igual que sucede con las conexiones, solo están presentes si existe topología

clásico, en el cual podemos conocer la forma de un área dada o el recorrido que sigue una determinada carretera, pero no las relaciones existentes. Únicamente disponemos del trazo con el que se han dibujado estos elementos. Por esta razón, y como se ha dicho, un modelo vectorial sin topología es perfectamente válido para la representación de cualquier tipo de información en formato vectorial, pero no tanto para su análisis.

El almacenamiento de entidades basado en una mera lista de coordenadas de cada entidad se conoce popularmente como *spaghetti*, pues si pensamos en una capa de líneas sin topología que se entrecruzan en el espacio, esta se asemejan en cierta forma a un caótico plato de *spaguettis* sin orden ni relación entre ellos.

La mayor ventaja de este modelo es su simplicidad, razón por la cual es la habitual en muchos de los SIG más populares. Para muchos usuarios, es suficiente trabajar con datos vectoriales sin topología, pues las labores frecuentes que desarrollan, tales como consultas (capítulo 12) o creación de mapas derivados, no requiere conocer las relaciones topológicas existentes.

Gran parte de las operaciones que se desarrollan en un SIG no requieren topología, y por ello no es necesario asumir siempre el coste que implica trabajar con ella (mayor complejidad en general). Una gran parte de la información vectorial disponible no incluye topología, ya que o bien esta no se incorporó en el momento de la digitalización, o bien el formato de fichero en el que está almacenada no soporta su almacenamiento.

En otros casos, la propia naturaleza de la variable que recogemos puede requerir almacenarse sin topología, o bien puede ser que no existan relaciones topológicas que representar. En una capa de polígonos en la cual se recojan las áreas de influencia de unos determinado fenómenos puntuales, esas áreas pueden perfectamente solaparse. No existe en este caso esa relación que hace que el conjunto de polígonos que las representan cubra la totalidad del espacio y cada punto pertenezca a una sola entidad. En este caso, un punto puede estar afectado por uno, varios o ninguno de dichos fenómenos puntuales, y por tanto pertenecer a una, varias o ninguna de las entidades poligonales que representan sus respectivas áreas de afección. Al modificar una de ellas (por ejemplo, si el fenómeno puntual que la origina varía su intensidad), las demás geometrías no deberían verse afectadas. No existe como tal una relación que deba recogerse en forma de topología.

Con topología

La alternativa al modelo vectorial sin topología es el almacenamiento explícito de las relaciones topológicas. Si se recoge de forma individual toda la información espacial correspondiente a cada entidad, la topología se pierde, pues no se considera al conjunto de entidades como un conjunto en el cual existen relaciones internas, sino como una simple colección de cosas. Para recoger la topología es necesario considerar todos los puntos que constituyen las entidades, y después *formar* las entidades a partir de ese todo de puntos, considerando en el proceso que un mismo punto puede pertenecer a varias entidades. Esto es lo que se denomina frecuentemente un *diccionario de puntos*, ya que contiene las definiciones de estos (sus coordenadas) y en base a ellos se construyen las distintas geometrías.

Esta forma de considerar el conjunto de entidades evita, además, la redundancia en los datos. Por ejemplo, para el caso mostrado en la figura 5.11, y en caso de no tener topología, el punto que es movido se encuentra almacenado dos veces, una por cada polígono. Al desplazarlo, solo se modifica una copia de dicha coordenada, la que pertenece al polígono editado, mientras que la otra permanece en su lugar. Si se dispone de topología, este punto se almacena una única vez, y al desplazarse se modifican las fronteras de todos los elementos (líneas o polígonos, según el caso) cuya frontera incluye dicho punto.

La denominación de *diccionario de puntos* que se mencionaba anteriormente es muy reveladora en este sentido. Si los puntos son como las palabras de un diccionario y los polígonos como frases o párrafos, basta pensar en lo poco práctico que sería escribir una frase en la que debiera definirse cada palabra al introducirla en dicha frase. Resulta mucho más adecuado (y ahorra esfuerzos al escritor), utilizar las palabras simplemente, y después definir estas en un diccionario en caso de que el lector no las conozca y necesite una referencia. Con el caso de los puntos sucede algo similar.

Existen diversos modelos para almacenar tanto las propias geometrías como sus relaciones inherentes, dos de los cuales se muestran en la figura 5.13 mediante sendos ejemplos en los que se codifican polígonos y líneas.

El primero de estos modelos es un modelo de carácter histórico denominado DIME (*Dual Independent Map Encoding*), desarrollado originalmente por el *US Bureau of the Census*, y posteriormente mejorado en el modelo TIGER, empleado para la digitalización de cartografía urbana. El segundo es el modelo *arco–nodo*, probablemente el más difundido y popular en la actualidad, aunque a este respecto los planteamientos existentes son muy variados.

En este modelo existen dos unidades fundamentales: Los nodos, que son puntos donde se *conectan* varias líneas; y los *arcos*, que son líneas entre dos nodos. Estas líneas no han de ser rectas, ya que pueden contener en su recorrido *vértices*. Los vértices son en realidad los puntos que solo pertenecen a una entidad, mientras que los nodos pertenecen a varias de ellas.

Una capa de líneas se describe como un conjunto de arcos y nodos, de forma que, atendiendo a los nodos como enlaces entre las líneas, se pueden conocer las relaciones entre ellas. En el caso de polígonos, estos se forman con conjuntos de arcos que delimitan las fronteras. Los polígonos que son adyacentes comparten uno o más arcos, quedando establecida así mediante ellos la relación topológica.

En el caso del modelo DIME, sin embargo, vemos que cada línea recta entre dos puntos se trata como una unidad, es decir, que todos los vértices son considerados como nodos y los arcos se componen siempre de una sola línea. El arco es en realidad un segmento. En ambos casos,

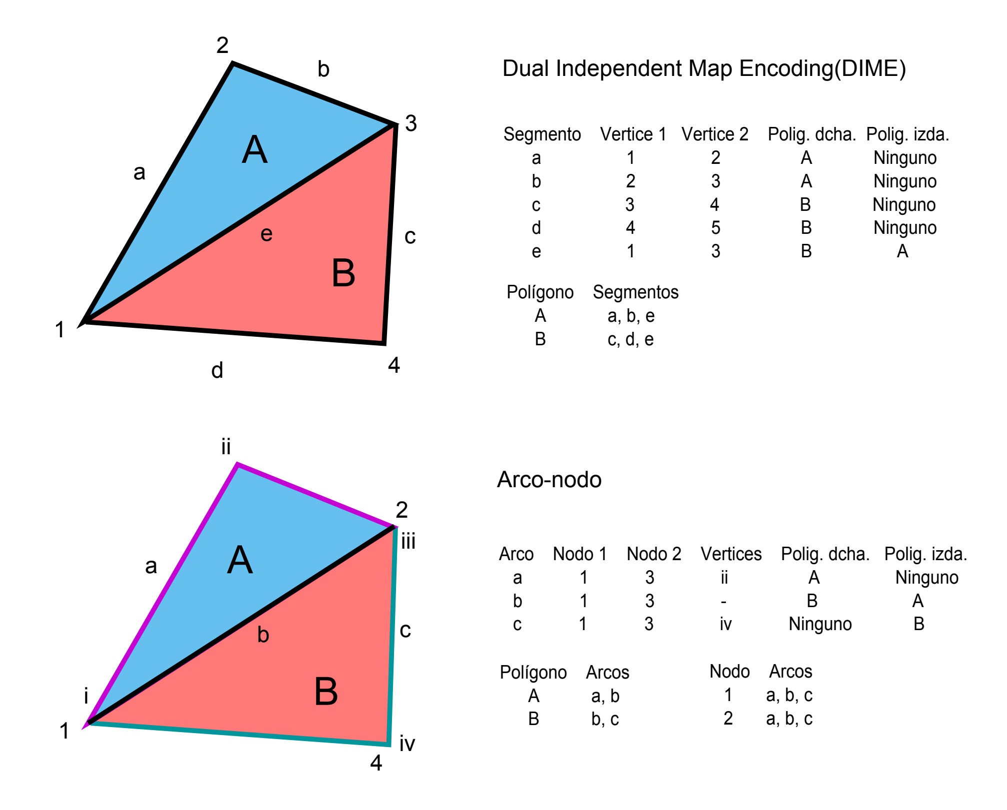


Figura 5.13: Dos modelos para representar la topología de líneas y polígonos. a) DIME, b) arco–nodo.

no obstante, cada arco tiene un inicio y un final —y por tanto una dirección—, y puede definirse un lado derecho y otro izquierdo según se avanza en dicha dirección. Como puede verse, también en ambos modelos se recoge explícitamente qué polígono, en caso de haber alguno, se sitúa a cada lado del arco.

La información que se recoge según estos modelos, vemos que se divide en bloques según los distintos niveles, desde los puntos, que han de recogerse en un diccionario de puntos (aunque este no queda reflejado en las tablas de la figura), pasando por los segmentos o arcos, y hasta los polígonos, definidos estos en base a los anteriores.

Con independencia del modelo, y sin entrar en más detalles, todos estos elementos en conjunto sirven para recoger las relaciones existentes entre los elementos, de tal modo que pueden llevarse a cabo también aquellas operaciones que no dependen exclusivamente de la posición, sino asimismo de otra serie de propiedades.

Dentro de los modelos existentes, encontramos asimismo variaciones en función de la tarea principal que se desee realizar. La eficiencia de cierto tipo de cálculos puede aumentarse notablemente si se elige un modelo de representación óptimo, como podemos ver si analizamos una de las operaciones más comunes: el cálculo de rutas óptimas entre dos puntos (los detalles sobre este cálculo se exponen en el capítulo 20, aquí

por el momento únicamente mostraremos sus implicaciones en los modelos de representación).

Para calcular la ruta óptima entre dos puntos dados de una red necesitamos conocer qué nodos de la red están conectados entre sí y por qué vías están conectados, ya que las características de estas condicionan el movimiento. La información necesaria para este cálculo puede almacenarse perfectamente según un modelo arco–nodo como el que ya conocemos, pero considerando las particularidades del análisis que queremos realizar, existen otros modelos más apropiados.

Por ejemplo, se puede tener en cuenta que los vértices de un nodo no tienen relevancia alguna. Si el tránsito se realiza entre dos nodos, a efectos del cálculo es indiferente que el tramo que los une tenga unos u otros vértices. Lo único que importa es saber que existe un tramo que los conecta y las características de ese tramo como, por ejemplo, el tiempo que cuesta recorrerlo o si conecta el nodo A con el B y el B con el A o solo lo hace en una de las direcciones anteriores. Por ello, en el caso del análisis de redes, la clave reside en almacenar de forma eficiente los nodos y las relaciones, pues estos son los elementos esenciales para efectuar los cálculos

Algunos modelos empleados comúnmente para el almacenamiento de redes son los siguientes [28]:

- Matriz de incidencias arco–nodo
- Matriz de adyacencias nodo–nodo
- Listas de adyacencia
- Estrella directa e inversa[5]

La matriz de adyacencias nodo–nodo es sumamente sencilla, ya que simplemente, para un número n de nodos, contiene una matriz de tamaño $n \times n$, en la que cada elemento (i, j) indica la existencia o no de conexión entre los nodos i y j y la naturaleza de dicha conexión. Si el elemento es igual a cero indica que no existe posibilidad de desplazarse directamente del nodo i al nodo j. En caso contrario, el valor es igual a la propiedad que se desee recoger del tramo, por ejemplo el tiempo que se tarda en recorrer o la velocidad máxima a la que puede hacerse ese recorrido.

La gran ventaja de este método es su gran sencillez, que deriva en sencillas implementaciones de los algoritmos correspondientes.

El método de estrella directa e inversa, por su parte, no es tan sencillo (una descripción algo más detallada puede encontrarse en [28]), pero es el más eficaz [32], y sus tiempos de cálculo asociados son los menores de entre todos los anteriores.

Más allá de los detalles particulares del modelo de representación, lo importante es tener presente que existen diversas formas de representar el dato geográfico, y que cada una de ellas tiene sus ventajas e inconvenientes en relación con la función que los datos hayan de desempeñar.

TIN

Hemos visto cómo una capa vectorial con topología nos sirve para modelizar ventajosamente elementos como una red de vías o una teselación del espacio en, por ejemplo, diferentes clases de usos de suelo. Además de esto, la incorporación de topología sirve para mejorar la representación de campos mediante modelos vectoriales, permitiendo la aparición de modelos como los TIN, ya presentados con anterioridad.

Un TIN [33] es una red formada por un conjunto de triángulos interconectados, cada uno de los cuales representa a una zona de características homogéneas en lo que a la variable estudiada respecta. Debido a esto, y como puede verse en

[5]Forward and reverse star

la figura 5.1, el número de triángulos varía según las características propias de la zona.

En aquellos lugares en los que se dé una gran variación (en caso de recoger el relieve será en las áreas más abruptas), se utiliza un gran número de triángulos para recoger toda esa variabilidad. Cuando, por el contrario, los valores no varían de forma tan notable (zonas de relieve más llano), pueden emplearse menos triángulos. Puesto que cada triángulo está formado, como todo polígono, por puntos, podemos decir que se necesitan menos puntos para almacenar un terreno si este es llano que si este es muy abrupto.

Cada triángulo tienen unas propiedades constantes, como corresponde al modelo vectorial. En particular, se considera habitualmente que todos los puntos dentro de un mismo triángulo constituyen un plano, con una pendiente y una orientación fija por tanto.

La topología del modelo permite llevar a cabo análisis diversos sobre un TIN, ya que para cada triángulo se tiene conocimiento de cuáles son los adyacentes a este, y es en el análisis de dichos adyacentes en el que se basan gran parte de los algoritmos. Este análisis resulta sencillo de implementar en una capa ráster, pues su propia estructura informa directamente de las celdas circundantes, pero en el caso vectorial requiere la presencia de topología para plantear un esquema similar de operación.

El análisis de los TIN no se desarrolla en detalle en este libro, pero resulta interesante recalcar en este punto que resulta posible de igual modo, y ello es debido a la presencia de topología en la propia estructura del modelo de representación.

Las particularidades del TIN hacen que existan sub–modelos principales para almacenar el conjunto de triángulos, distintos del habitual arco–nodo, y pensados específicamente para responder a las necesidades que los TIN demandan como modelos vectoriales para representar variables continuas (en este sentido, es algo muy similar al caso que veíamos anteriormente de las redes). Estos modelos son dos, principalmente:

- Almacenamiento de los triángulos uno por uno, cada uno con las coordenadas de todos sus tres puntos (coordenadas tridimensionales, no planas) y un código de identificación, y almacenamiento de los códigos de los triángulos adyacentes.
- Almacenamiento de los vértices y un código para cada uno de ellos, así como los códigos de los vértices a los que se encuentra conectado, en un orden establecido (horario o antihorario).

Más información sobre TIN puede encontrarse en [34]. La creación de TIN se trata con más detalle en el capítulo 18.

5.3.3. Raster *vs* vectorial

Resulta obvio que las diferencias entre los modelos ráster y vectorial son muy notables, y que cada uno de ellos posee sus propias ventajas e inconvenientes. Desde los primeros tiempos de los SIG, ha existido una clara tendencia a separar ambas realidades en la implementación, de tal modo que los primeros SIG manejaban datos en formato ráster o bien en formato vectorial, pero no ambos. En cierta medida, parecía existir un conflicto entre ambos modelos, el cual ha perdurado aún hoy en algunos conceptos. Con el paso del tiempo, la separación ráster–vectorial ha cambiado notablemente, y ha quedado claro que un SIG eficaz debe ser capaz de manejar todo tipo datos geográficos con independencia del modelo de datos empleado.

La comparación entre ambos modelos resulta necesaria para hacer un uso correcto de ellos, eligiendo en cada caso el más adecuado, y combinándolos de la manera óptima. Algunos aspectos a los cuales puede atenderse para comparar uno y otro modelo son los siguientes:

- **Planteamiento**. Íntimamente ligados con los modelos conceptuales del espacio geográfico, los planteamientos de los modelos de representación ráster y vectorial son diferentes en su naturaleza. El modelo ráster hace más énfasis en aquella característica del espacio que analizamos (*qué* y *cómo*), mientras que el modelo vectorial da prioridad a la localización de dicha característica (*dónde*)
- **Precisión**. El modelo ráster tiene su precisión limitada por el tamaño de celda. Las entidades menores que dicho tamaño de celda no pueden recogerse, y la variación espacial que sucede dentro del espacio de la celda tampoco.

  Asimismo, existe una imprecisión en las formas. El detalle con el que puede recogerse la forma de una entidad geográfica según el modelo vectorial es, en la práctica, ilimitado, mientras que, como puede verse en la imagen 5.14, el modelo ráster restringe las formas a ángulos rectos, ya que la unidad base es un cuadrado.

  El perímetro de una entidad geográfica estará compuesto por líneas horizontales o verticales exclusivamente y, además, su longitud y la superficie que encierra serán respectivamente múltiplos del tamaño de celda y el área de dicha celda. Esta es la principal razón por la cual, si el uso principal que se le va a dar a una capa es su representación gráfica, deba optarse por el modelo vectorial. En caso contrario, y salvo que la resolución sea suficientemente alta, los mapas creados mostraran la falta de resolución y podrán distinguirse las unidades mínimas de la capas ráster (al igual que pasa en una imagen digital *pixelada*), teniendo un aspecto que no es el propio de un mapa, tal y como estamos acostumbrados a usarlo.

  El hecho de que dentro de una celda el valor de la variable recogida sea constante, da lugar a ambigüedades como la mostrada en la figura 5.15, donde una celda está ocupada por dos valores distintos, pero solo puede asignársele uno de ellos, debiendo establecerse algún criterio sistemático para llevar esto a cabo.

  Un hecho similar sucede en el ejemplo de la capa de vías. Algunas celdas son atravesadas por más de una vía, pero esa información se pierde, ya que el tamaño de celda no es suficiente para recogerla. La celda en cuestión aparece como celda de vía, pero no sabemos cuántas diferentes la atraviesan, ni tampoco si entre ellas están enlazadas o no.

  Aun si el modelo vectorial puede capturar elementos geográficos con mayor precisión, hay que tener en cuenta que esa precisión es de la representación como tal, es decir, del modelo en sí, pero no del dato que tenemos en dicho formato vectorial. La precisión de este dependerá de otros condicionantes tales como la escala de trabajo. Existe siempre incertidumbre en los datos, y el modelo de almacenamiento no excluye esta circunstancia. Los aspectos relativos a la calidad de los datos, tanto para datos ráster como vectoriales, se desarrollan en profundidad en el capítulo 7.
- **Volumen de almacenamiento**. El número de elementos a almacenar es, en general, muy superior en el caso del modelo ráster. Esto es así debido a que toda la superficie a recoger se divide en las mismas unidades, independientemente de la complejidad de la variable en cada punto o de la necesidad de estudiarla con mayor o menor detalle en unos puntos que en otros. Para variables que se conciban mejor según un modelo conceptual de entidades discretas, el modelo vectorial resulta más adecuado, ya que no es necesario registrar de modo explicito las zonas sin entidades, mientras que en el modelo ráster estas deben registrarse de igual modo que aquellas en las que sí existe información relevante.

  Los modelos de almacenamiento ráster que veremos en el siguiente punto solucionan en parte

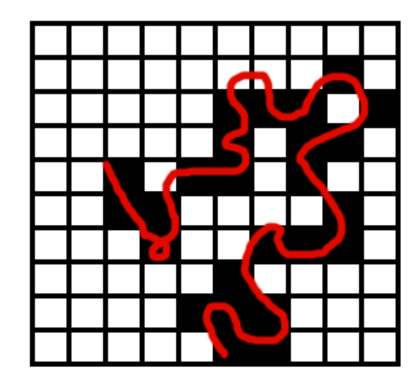


Figura 5.14: Imprecisión de forma en el modelo de representación ráster. La división del espacio en unidades cuadradas impide la representación fiel de entidades tales como curvas como la mostrada en trazo rojo en la figura.

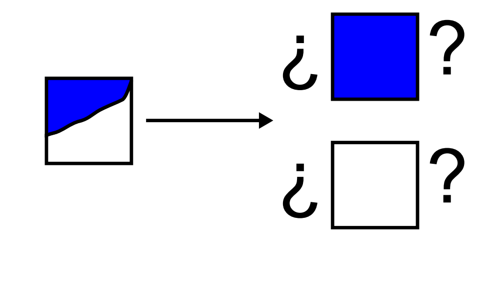


Figura 5.15: Ambigüedad en la asignación de valores a una celda en una capa ráster, debido al tamaño de esta, que condiciona la precisión con la que puede recogerse la realidad existente sobre el terreno.

el problema de los grandes volúmenes de datos, y son un elemento importante para una implementación eficiente.

- **Complejidad**. La regularidad y sistematicidad de las mallas ráster hacen sencillo el implementar algoritmos de análisis, muy especialmente aquellos que implican el uso combinado de varias capas. Cuando estas capas están en formato ráster y existe coincidencia entre sus mallas de celdas, el análisis conjunto de estas resulta inmediato. Por el contrario, la irregularidad espacial de las capas vectoriales hace que la implementación de los mismos algoritmos sea sumamente más compleja si se trabaja con estas capas.

  La sencillez de las capas ráster, tanto en su concepto como en su implementación, se ve apoyada además por el hecho de que una capa ráster se puede asemejar a una matriz, y por tanto aplicar sobre ella una serie de herramientas y elementos matemáticos en muchos casos bien conocidos y de fácil comprensión.

  Existe de igual forma una complejidad distinta en términos de proceso y cálculo. Los algoritmos sobre una base ráster pueden ser costosos por la necesidad de aplicarlos sobre un número elevado de celdas y un gran volumen de datos (véase el punto anterior). Por el contrario, los algoritmos sobre una base vectorial son costosos debido a que las operaciones matemáticas que implican son más complejas y requieren mayor número de cálculos (aunque los volúmenes manejados puedan también ser notables).

Mas allá de las anteriores diferencias, a la hora de planificar un trabajo dentro de un SIG y elegir los datos que emplearemos y el modelo de representación ideal, lo importante es entender que no existe un modelo de representación idóneo de forma global, sino que esta idoneidad depende de muchos factores, como por ejemplo:

- **Tipo de variable o fenómeno a recoger**. Como ya sabemos, algunas variables, en función de su variabilidad y comportamiento espacial, son más adecuadas para el modelo vectorial, mientras que otras lo son para el modelo ráster. Por ejemplo, en el caso de variables que requieran una intensidad de muestreo distinta según la localización (variables que resulta interesante estudiar con más detalle en unos puntos que en

otros) puede resultar más lógico recogerlas de forma vectorial, pues el modelo ráster implica una intensidad de muestreo constante a lo largo del área estudiada.

- **Tipo de análisis o tarea a realizar sobre dicha variable**. El uso que demos a una capa temática condiciona en gran medida el modelo de datos idóneo. Por ejemplo en el caso de una capa de elevaciones, su análisis se lleva mejor a cabo si esta información está recogida según el modelo ráster. Sin embargo, si el objetivo principal es la visualización de esa elevación en conjunto con otras variables, unas curvas de nivel pueden resultar más adecuadas, ya que, entre otras cosas, no interfieren tanto con otros elementos a la hora de diseñar un mapa con todas esas variables.
- **Contexto de trabajo**. Por ejemplo, si queremos trabajar con imágenes, esto nos condiciona al empleo de datos ráster, ya que resulta mucho más sencillo combinarlos con las imágenes, las cuales siempre se presentan como capas ráster.

Así, en el desarrollo de un trabajo pueden aparecer circunstancias que hagan más adecuado utilizar el modelo ráster y otras en las que el modelo vectorial sea más idóneo. En tal caso, deben combinarse ambas, pues es de esta forma como se obtendrán mejores resultados. Un usuario de SIG no debe limitarse a trabajar de forma general con un único modelo de datos, con independencia del tipo de tarea que desempeñe, pues en cualquier caso ambos pueden aportar alguna ventaja.

Por último, es importante tener en cuenta que existen procedimientos para convertir entre los formatos ráster y vectorial, de forma que el disponer de datos en un modelo de representación particular no implica que debamos desarrollar nuestro trabajo sobre dichos datos directamente, sino que podemos efectuar previamente una conversión. Los capítulos 14 y 18 tratan estos temas en profundidad.

5.4. Modelos de almacenamiento

Los modelos de almacenamiento son el ultimo escalón en la cadena de etapas distintas que llevan desde la realidad existente al conjunto de simples valores numéricos que almacenamos y manejamos en un SIG y que modelizan dicha realidad. Los modelos de representación definen una forma de recoger la realidad mediante unidades básicas (sean estas celdas en una malla, o bien primitivas geométricas definidas de una u otra manera), mientras que los modelos de almacenamiento plantean básicamente un esquema de cómo convertir dichas unidades en valores numéricos de la forma más eficiente. Es decir, cómo *escribir* dichos valores en un soporte digital o guardarlos en la memoria del ordenador de la mejor manera posible.

Los modelos de almacenamiento deben atender principalmente a dos necesidades básicas, que son las que definirán su idoneidad para cada tarea y tipo de dato:

- Minimizar el espacio ocupado por los datos.
- Maximizar la eficiencia de cálculo.

La primera necesidad es especialmente importante en los datos raster, pues, como ya se ha dicho, estos son con frecuencia muy voluminosos. Un modelo de representación que minimice el tamaño de los datos, unido a un manejo óptimo de memoria, son requisitos de suma importancia para todo SIG que trabaje con capas ráster.

La necesidad de maximizar la eficiencia de cálculo afecta principalmente a las representaciones vectoriales, ya que en ellas las operaciones son complejas. La forma en que se estructuran los valores de cada entidad ha de minimizar el numero de accesos necesarios a estos, para de este modo obtener un mejor rendimiento en todas las operaciones de análisis.

5.4.1. Modelos para representaciones ráster

El principal problema relativo al almacenamiento de capas ráster se presenta para el conjunto de valores de las distintas celdas, que constituye la parte más voluminosa de la información recogida. Las coordenadas de las celdas de referencia o el tamaño de celda, por su escaso volumen, no conllevan dificultad alguna, y es en el almacenamiento de la malla de celdas en sí donde se encuentran las diferencias entre unos y otros modelos.

La forma más inmediata de almacenar una capa ráster es almacenar simplemente sus valores uno a uno, en una estructura similar a la que la propia capa representa. Para el caso más habitual de capas con celdas cuadradas, sabemos que la malla de datos correspondiente se puede asimilar a una matriz, con las implicaciones que esto tiene a la hora de su manejo. Así, la forma más directa de recoger una malla de datos ráster es mediante una matriz de datos. Esta forma de almacenamiento tiene las siguiente ventajas [35]:

- **Formato muy intuitivo**. La mayoría de desarrolladores está familiarizado con el concepto de matriz y con las operaciones de calculo matricial que pueden aplicarse sobre estas.
- **Sencillez en la implementación**. Los lenguajes de programación soportan sin problemas el uso de matrices bidimensionales y una serie de operaciones básicas sobre ellas.
- **Estructura**. Las mismas operaciones pueden aplicarse sobre todos los valores de la matriz de igual modo (todas las posiciones de la matriz son *iguales* desde este punto de vista), lo que simplifica la implementación de operaciones.
- **Iterabilidad**. Resulta igualmente sencillo recorrer la matriz e iterar sobre la misma, lo cual refuerza lo anterior y simplifica aún más la implementación de todo tipo de procesos.

No obstante, el almacenamiento de todos los valores de forma idéntica ignora el hecho de que pueden existir valores similares en zonas concretas, que pueden recogerse de formas mucho mejores que una serie de números iguales. En otras palabras, y de modo similar a como ocurre con el propio modelo de representación ráster, la estructura regular que confiere las ventajas es también la responsable de la mayor parte de los inconvenientes.

Como veremos en el capítulo 11, las zonas próximas entre sí (es decir, en el caso de una capa ráster, las celdas próximas entre sí), tienden a tener valores similares, en lo que se conoce como *autocorrelación espacial*. No considerar este hecho lleva al almacenamiento de información redundante, y ese es precisamente el principal problema del almacenamiento directo de una capa ráster mediante una matriz. Almacenando explícitamente todos los valores de la malla se desperdicia en muchos casos una gran cantidad de espacio (sea este en memoria, disco u otro soporte cualquiera).

Podemos ver dos ejemplos claros de esto en las figuras 5.2 y 5.6. En la primera, existen únicamente dos valores: los correspondientes a las celdas sobre las que se sitúa una vía, o los correspondientes a las celdas donde estas no aparecen. Estos últimos ocupan la gran mayor parte de la capa, y lo hacen en bloque, de tal forma que almacenándolos individualmente se acaba teniendo una matriz de datos donde la practica totalidad de ellos son idénticos. Como es fácil de entender, este forma de proceder no es la más adecuada, al menos en términos de volumen de almacenamiento.

En la segunda imagen, las zonas que aparecen como consecuencia de la rotación de la imagen no contienen datos (esto es, contendrán el valor arbitrario que codifica la falta de datos). Estas zonas también constituyen grandes bloques de

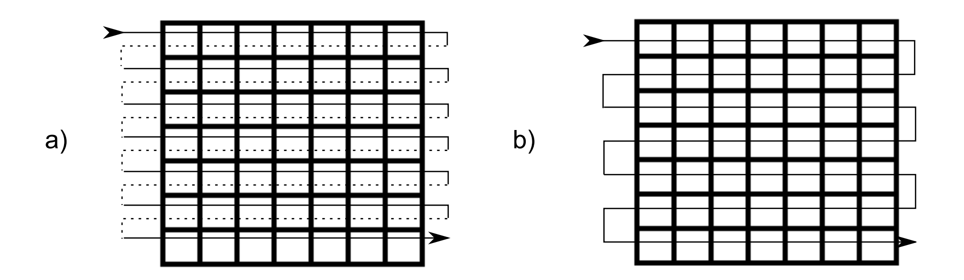


Figura 5.16: Ordenes de escaneo. a) fila a fila sin retorno, b) fila a fila con retorno.

celdas contiguas, con lo que el almacenamiento de todos los valores también es una solución altamente redundante, especialmente en estas zonas fuera de la imagen como tal.

La solución más habitual para considerar la redundancia de valores y lograr una compresión eficaz de los datos es la técnica denominada *Run–Length Encoding*. Esta técnica sencilla codifica una serie de n valores idénticos como un par de valores, el primero de los cuales representa el valor que se repite n veces, y el segundo es el número de veces que se repite, esto es, n.

Así, si la primera fila de la capa de vías en formato ráster no aparece ninguna celda de vía, todas las celdas de dicha fila contendrán el valor con que se codifica la ausencia de estas (sea, por ejemplo, el valor 0). El almacenamiento directo de todos los valores de la fila requeriría tantos valores como columnas existan (sea n el ancho de la fila), mientras que utilizando *Run–Length Encoding*, bastaría con almacenar el par (0, n).

A la hora de tratar el conjunto de todas las celdas, se define un orden en el que recorrerla, denominado *orden de barrido* o *de escaneo* (Figura 5.16), de tal modo que la matriz bidimensional queda reducida a una cadena de valores, es decir, a un vector unidimensional. Los distintos *trozos* de esa cadena se van codificando según el esquema anterior, de tal modo que cuando aparecen muchos valores iguales consecutivos, estos pueden sustituirse ventajosamente por un único par de valores.

La forma más sencilla de recorrer la imagen es hacerlo por filas, empezando por la fila superior y desplazándose de derecha a izquierda (Figura 5.16a). No obstante, el salto que se produce al final de cada fila suele implicar una discontinuidad en los valores. Invirtiendo la dirección del recorrido en cada fila, se tiene el orden mostrado en la figura 5.16b, el cual suele tener como resultado mayores niveles de compresión de datos, ya que la cadena resultante de recorrer la imagen contiene *trozos* generalmente de mayor tamaño.

Un esquema de barrido más complejo es el basado en el denominado *orden de Morton* [36]. El orden de Morton (también conocido como *orden Z*), se basa en una curva de carácter recursivo que recorre las celdas de la matriz siguiendo tramos en forma de Z, de ahí el nombre. En la primera iteración se divide el conjunto de celdas en cuatro bloques, los cuales se recorren siguiendo el antedicho recorrido en Z. Si los bloques contienen a su vez más de una celda, se siguen subdividiendo a su vez de forma idéntica, y así hasta que no pueda continuarse este proceso.

La matriz que contiene los valores de orden de Morton (el orden en que se visita cada celda según el esquema anterior), se conoce como *Matriz de Morton*, la cual ya citamos por su importancia histórica en el capítulo 2.

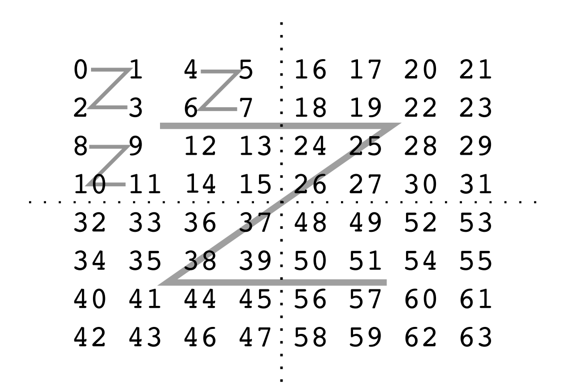


Figura 5.17: Un ejemplo sencillo de barrido de una capa ráster según órdenes de Morton. Los valores en las celdas no indican los valores de la variable, sino el orden en que se visita dicha celda según este esquema de barrido

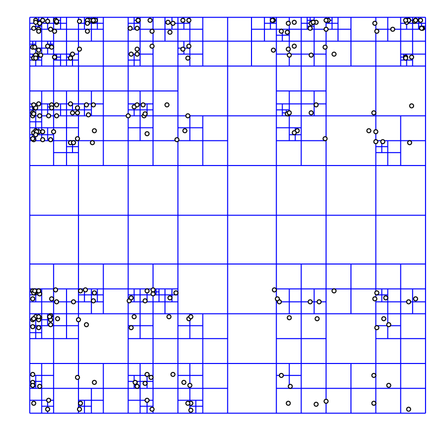


Figura 5.18: Ejemplo de un árbol cuaternario. En las zonas con más variabilidad (mayor densidad de puntos), los cuadrantes se subdividen hasta una profundidad mayor. La estructura es tal que cada cuadrante tiene dentro a lo sumo un punto. (Tomado de Wikipedia)

Un ejemplo de este orden de barrido aplicado a una pequeña matriz puede verse en la figura 5.17.

Una estructura más avanzada son los denominados *Quadtrees* o árboles cuaternarios. Estas estructuras también dividen el espacio en cuadrantes sucesivamente, pero lo hacen con más profundidad en aquellas zonas que así lo requieran por contener mayor número de elementos y necesitar mayor resolución. En el caso de una capa ráster, se requerirá más detalle siempre que todas las celdas dentro de un cuadrante no tengan el mismo valor. En el caso más extremo, se ha de descender hasta el nivel de una sola celda, pero puede ser que un bloque de celdas contiguas tenga el mismo valor, en cuyo caso el cuadrante correspondiente las engloba a todas y las define con dicho único valor, sin necesidad de subdividirse más. De este modo, se adapta el modelo de almacenamiento a la propia estructura de la capa y al comportamiento que en esta muestra la variable estudiada.

Un ejemplo gráfico de un árbol cuaternario puede encontrarse en la figura 5.18. Los arboles cuaternarios son empleados también en los *índices espaciales*, asociados a representaciones vectoriales, que veremos en 12.4.2 (de hecho, puede apreciarse que la figura anterior representa la aplicación de un árbol cuaternario a un conjunto de puntos, no a una capa ráster, aunque el concepto es el mismo y su aplicación a este segundo caso se realiza como ya se ha mencionado previamente).

Los quadtrees son estructuras complejas, y no profundizaremos más en su descripción dentro de este capítulo. Para el lector interesado, la definición original de esta estructura de datos puede encontrarse en [37].

Es importante reseñar que cuando la capa ráster contiene una información tal como una red viaria, la cual es susceptible de presentar valores idénticos en celdas contiguas, la codificación de tipo *Run–Length* —con cualquiera de los esquemas de barrido anteriores— es ventajosa. Sin

embargo, no lo es tanto cuando se trabaja con otro tipo de variables.

En una capa con valores de elevación, las celdas próximas tendrán valores parecidos pero no idénticos, con lo que no podrá sacarse partido a esta forma de almacenamiento. Más aún, en estos casos el volumen ocupado por los datos no solo no disminuye, sino que aumenta. Es por ello que los SIG han de implementar igualmente la capacidad de poder trabajar con uno u otro modelo de almacenamiento según los casos, bien sea por elección directa del usuario o tomándose de forma automática el que el propio sistema considere más adecuado en cada ocasión.

Aunque el mayor problema de las capas ráster es su gran volumen, también existen diversas alternativas enfocadas a mejorar la velocidad de acceso a datos y el rendimiento de las operaciones sobre estas capas. Estas alternativas afectan a las imágenes con múltiples bandas, ya que estas, como dijimos, se recogen en un único fichero, en el cual se incorpora toda la información de las distintas bandas.

La forma en la que las bandas se tratan dentro del fichero y el modo en que se ordenan los píxeles de las distintas bandas, ambas definen el esquema de almacenamiento, presentando cada uno de ellos una serie de ventajas de rendimiento en función de la actividad principal que se vaya a desarrollar con la imagen. Tres son los esquemas principales:

- *Band Sequential* (BSQ). Los valores se almacenan por bandas. Es decir, primero todos los píxeles de la banda 1, después los de la banda 2, y así sucesivamente. Este tipo de esquema da prioridad a la componente espacial, ya que permite acceder rápidamente a todo el espacio cubierto por una banda, puesto que los píxeles de dicha banda se encuentran almacenados en posiciones contiguas.
- *Band Interleaved by Pixel* (BIP). Los valores se almacenan ordenados por posiciones de píxel. Es decir, primero se almacenan todos los valores correspondientes al píxel (0, 0)[6] (en todas las bandas existentes), después los correspondientes al (0,1)[7], y así sucesivamente. En caso de que lo que interese sea, para un píxel dado, conocer toda la información disponible (su valor en todas las bandas), el esquema BIP es más ventajoso, ya que permite accesos rápidos a este tipo de información, sin necesidad de «saltar» de un valor a otro como sucedería en el caso del esquema BSQ. A la hora del acceso a los datos, se prima la información espectral sobre la espacial.
- *Band Interleaved by Lines* (BIL). Es un esquema intermedio en el que se recogen los valores por filas. Esto es, primero la fila 1 de la banda 1, luego la de la banda 2, y así sucesivamente. Posteriormente se recoge la fila 2 para todas las bandas, y de este modo hasta cubrir toda la imagen. Se trata de un esquema intermedio entre los anteriores, permitiendo un acceso rápido tanto a la información espacial como a la información espectral de las bandas.

La figura 5.19 se muestra un ejemplo muy sencillo de los anteriores esquemas. Para una imagen de 2×2 celdas y dos bandas, se recoge el orden en que se almacenaría cada valor según cada uno de dichos esquemas.

5.4.2. Modelos para representaciones vectoriales

Al igual que para el modelo ráster, existen para el modelo vectorial diferentes alternativas a la hora de almacenar los elementos que componen una capa. En realidad, ya hemos visto dentro de

[6]Es una habitual empezar a contar en cero en lugar de en uno las coordenadas fila/columna de una imagen
[7]Es habitual recorrer la imagen por filas, de forma que (0,1) representa la primera fila y la segunda columna

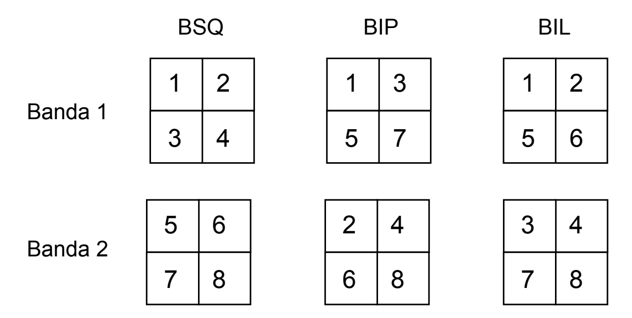


Figura 5.19: Esquemas de almacenamiento para imágenes multibanda. Los números indican el orden en que se almacena cada valor.

este capítulo algo que se asemeja a un modelo de almacenamiento, pues los modelo topológicos como DIME o el modelo *arco–nodo*, o los detallados para el caso particular de las redes, todos son en realidad esquemas de almacenamiento para el conjunto de *piezas* que componen esa estructura topológica que se quiere almacenar. No obstante, también tienen algo de modelos de representación, pues existe variación en la forma en que conciben las partes de cada entidad (arcos entre dos nodos con o sin vértices intermedios, según el modelo).

En realidad, la razón por la que se han presentado en una sección anterior es porque de ese modo ayudaban a comprender mejor la existencia o no de topología en una representación, y ese aspecto resulta más importante para el estudio de los SIG que los modelos de almacenamiento. Estos, como se ha dicho, están a un nivel más bajo y alejado del usuario.

En general, los modelos de datos vectoriales no buscan tanto la disminución de volumen de los datos como la obtención de una mayor eficacia en las operaciones y una simplificación de estas. Lógicamente, si los datos tienen un volumen menor, el tiempo que cualquier operación sobre ellos implica también ser menor. Aún así, la diferencia principal para este tipo de datos reside en la disminución de la complejidad en que estos se almacenan, disminuyendo las operaciones a realizar, así como la complejidad de la implementación de los correspondiente algoritmos (ambas habitualmente elevadas).

Para mejorar el rendimiento de las operaciones que trabajan con datos vectoriales, un factor clave es mejorar el acceso a los datos, de forma que, cuando se necesite acceder a unos datos concretos, estos puedan «encontrarse» de forma fácil. Por este motivo, un elemento importante en la representación de los datos vectoriales son los denominados *índices espaciales*.

El concepto de índice cuando se habla de datos es similar al concepto de índice referido a un libro como este. Aquí tienes un ejemplo muy sencillo para que lo comprendas mejor: si vas al principio de este libro, puedes ver su índice y saber dónde empieza este capítulo, de forma que si estas interesado en modelos relacionados con la información geográfica, sabes rápidamente que es en este bloque de páginas donde debes buscar lo que te interesa. Si no existiera ese índice, tendrías que ir revisando todas las páginas hasta que llegaras al principio del capítulo y te dieras cuenta de que aquí es donde está lo que buscas.

Este ejemplos muestra una situación similar a las que aparecen en el uso habitual de un SIG, en las cuales trabajamos sobre una parte del total de los datos. Igual que buscamos un capítulo, po-

demos querer, por ejemplo, todas las entidades de una capa que están en una zona particular del espacio. Disponer de un índice acelera el proceso de localizar esas entidades que nos interesan. Por trabajar con información espacial, tales índices se denominan índices espaciales.

Muchos de los procesos que veremos en la parte dedicada al análisis necesitan este tipo de índices para poder ejecutarse con un rendimiento adecuado. A medida que veamos estos procesos, se comprenderá mejor por qué la existencia de índices espaciales resulta necesaria e incluso imprescindible cuando disponemos de datos de gran volumen. En el capítulo 12 veremos información más detallada sobre la utilidad de los índices espaciales, ya que son vitales para la realización de consultas espaciales como las que se allí se explican.

Como ya hemos dicho, el objetivo de este tipo de estructuras para representar los datos espaciales no es disminuir el tamaño, sino mejorar el rendimiento de las operaciones sobre ellos. De hecho, y al contrario que en el caso de los modelos de representación ráster, en este caso no disminuye el espacio que ocupan los datos, sino todo lo contrario, ya que este aumenta. Un índice espacial es información adicional que incrementa la utilidad de dichos datos. Exactamente del mismo modo que el índice de este libro, que no sustituye al texto que ahora mismo estas leyendo, sino que se añade a este y te ayuda a manejarte a través de él y sacarle más partido.

La creación del índice espacial supone la creación de una estructura espacial en la cual se contienen objetos más simples que las propias entidades geométricas, estructuradas a su vez de forma también más sencilla que recogiendo sus coordenadas, y con un orden característico. Como hemos dicho, este índice espacial no sustituye al dato espacial, sino que lo complementa, optimizando la búsqueda de información dentro de este.

Existen dos enfoques principales para los índices espaciales: continuos y discretos [38]. Los continuos utilizan las coordenadas mismas de las entidades, simplificando la forma de estas, mientras que en los discretos la simplificación se aplica al espacio, discretizando este. En ambos, las entidades que se emplean son rectangulares en la mayoría de los casos. La figura 5.20 muestra la aproximación de una geometría poligonal que se obtiene en ambos tipos de modelos.

En el caso continuo, se sustituye toda la complejidad del polígono por simplemente cuatro puntos: aquellos que conforman el rectángulo dentro del que este se inscribe. En el caso discreto, se reduce el polígono a unas cuantas celdas de una malla. Realizar comprobaciones sobre estas estructuras resulta mucho más sencillo, y por ello se emplean para realizar aproximaciones que simplifican las operaciones[8].

Supongamos que utilizamos un índice espacial del primer tipo y queremos saber qué polígonos de una capa se intersecan con otro dado. Para ello, en primer lugar podemos comprobar los solapes existentes entre sus rectángulos. Si los rectángulos no se solapan, es inmediato ver que los polígonos tampoco, con lo que no es necesario ya operar con ellos. Ver si dos rectángulos se solapan es casi inmediato, mientras que esta misma operación para polígonos complejos requiere un numero mucho mayor de operaciones.

Debido al uso de rectángulos como elementos básicos, las estructuras que se emplean están específicamente diseñadas para contener o bien rectángulos (en el caso de entidades de líneas o de polígonos) o puntos (en el caso de entidades puntuales). Estas estructuras no son nuevas para nosotros, ya que hemos visto algunas de ellas en este mismo capítulo. Por ejemplo, para el caso

[8]Este proceso, conocido como *filtrado y refinamiento*, lo veremos en detalle en el capítulo 12

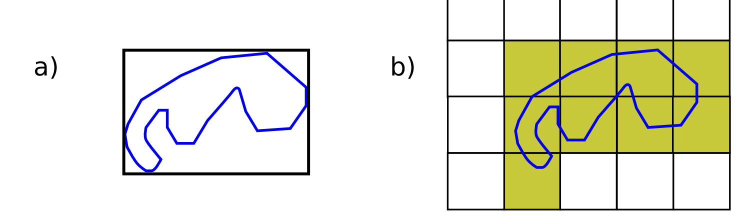


Figura 5.20: Aproximación continua (a) y discreta (b) para un índice espacial.

de una aproximación continua sobre una capa de puntos, los arboles cuaternarios (*quadtrees*) son una estructura de datos adecuada. Esta aplicación ya la vimos, de hecho, en la figura 5.18.

Es fácil advetir que los enfoques continuo y discreto se corresponden con las ideas correspondientes a los modelos de datos ráster y vectorial (aunque los índices espaciales de los que estamos hablando son para capas vectoriales). Es por ello que las estructuras que hemos visto para el almacenamiento de datos ráster pueden utilizarse también para recoger las distintas celdas de un índice espacial discreto. Así, la división en celdas hace necesario un orden de escaneo. El orden de Morton que ya conocemos se aplica en este caso, entre otros.

Una vez más, las estructuras de datos de todos estos índices espaciales suponen un elemento demasiado especifico para los contenidos de este libro, por lo que no se profundizará en su teoría. Referencias como [39] aportan descripciones más extensas para el lector interesado.

En caso de querer profundizar en los aspectos más técnicos de la representación del dato geográfico en general, tanto en formato ráster como vectorial, [40] ofrece información muy extensa al respecto.

5.5. Resumen

El proceso de almacenar la realidad y reducirla a un conjunto de valores numéricos manejables por un ordenador implica tres etapas fundamentales: creación de un modelo conceptual, adopción de un modelo de representación y codificación del anterior según un modelo de almacenamiento. Estos procesos dan lugar a la creación de las denominada *capas*, unidades fundamentales de información dentro de un SIG.

Dos son los modelos conceptuales más importantes: campos y entidades discretas. Estos a su vez se identifican en líneas generales con los dos principales modelos de representación: ráster y vectorial.

En el modelo ráster el espacio se divide sistemáticamente en unidades mínimas denominadas celdas, habitualmente de forma cuadrada. En el modelo vectorial se almacenan las distintas entidades geográficas a través de las coordenadas de los puntos que las componen. El concepto de *topología* es importante en el modelo vectorial, y en función de la forma en que se recojan las coordenadas de cada entidad, se almacenará o no la información topológica. El modelo arco–nodo es el más habitual para representar la topología.

La ultima etapa es la que conlleva el almacenamiento de los modelos de representación, convirtiendo los elementos base de estos en valores numéricos manejables por el ordenador. Cada modelo de representación tiene sus particulares modelos de almacenamiento, los cuales tratan de maximizar el rendimiento de las operaciones realizadas sobre los datos espaciales, al tiempo que reducen el espacio que dichos datos ocupan.

Capítulo 6

Fuentes principales de datos espaciales

6.1. Introducción

El origen de los datos con los que trabajamos en un SIG puede ser muy variado y presentarse en formas diversas. La metodología seguida en la recolección de datos condiciona directamente la forma en que estos datos llegan a nosotros, y por tanto el uso que les podemos dar dentro de un SIG o las operaciones que debemos realizar con ellos.

No hace tanto tiempo, toda la información que se manejaba dentro de un SIG tenía su origen en un mapa en papel, el cual debía *prepararse* para adaptarse a la naturaleza propia del SIG. El desarrollo de los SIG ya había comenzado a dar sus frutos y exisstían los programas, pero eran necesarios datos para utilizarlos. Sin embargo, los datos geográficos de los que se disponía no se encontraban en formato digital, por lo que no eran adecuados para su uso dentro de un SIG.

Una tarea básica en esos tiempos era la digitalización de cartografía, es decir, convertir los datos geográficos en formato impreso en datos en formato digital que un SIG pudiera manejar. La disponibilidad de datos digitales era baja, pero, como resulta lógico pensar, sí que existía una gran cantidad de datos geográficos en otros formatos tales como mapas, cartas de navegación, fotografías aéreas, etc. La tecnología ha ido avanzando y ya se producen datos directamente en formato digital, considerando específicamente la existencia de los SIG como herramientas básicas de manejo de datos geográficos. No obstante, los datos en formato impreso, así como las técnicas que se emplearon en su creación, siguen siendo válidas, y sirven igualmente para crear datos geográficos que podemos emplear en un SIG.

Hoy en día, la situación es bien distinta a la de aquellos primeros tiempos, y puede afirmarse que los orígenes a partir de los cuales se generan los datos geográficos son muy diversos. Esto es así porque aunan técnicas recientes y más adaptadas al entorno de los SIG con métodos clásicos que no han perdido su vigencia y valor.

A lo largo de este capítulo veremos las principales técnicas existentes para la creación de datos geograficos en un formato apto para su uso en un SIG, centrándonos en los pormenores de proceso y las particularidades de los datos generados en cada caso. Para ello, veremos todo el conjunto de fuentes de las cuales pueden provenir los datos con los que trabajamos en un SIG, desde las más modernas hasta las más antiguas, así como las metodologías que permiten convertir las formas no digitales en datos aptos para su uso en dicho SIG. El objetivo es que, al final del capítulo, se conozcan con detalle todas las formas en las que los datos geográficos pueden presentarse, se entiendan estas completamente con independencia de su origen, y se sepan utilizar y combinar todas las fuentes de datos.

6.2. Datos digitales y datos analógicos

La principal diferencia que se presenta desde la aparición de los SIG es la necesidad de utilizar datos digitales. Un SIG implica una aplicación informática, y esta se alimenta en última instancia exclusivamente de datos digitales. Esta es la razón por la que debemos alimentar nuestro SIG con una serie de valores numéricos, y llegar a ellos a partir de la realidad que se pretende modelizar implica toda una serie de etapas, las cuales ya vimos con detalle en el capítulo 5

Gran parte de los datos geográficos que se producen actualmente son en formato digital. Otros, a pesar de producirse hoy en día, no lo son directamente. Y junto a estos tenemos, como ya sabemos, todos los datos (que no son pocos) generados con anterioridad y que se presentan en diversas formas. Pero si deseamos trabajar con ellos en un SIG, de un modo u otro todos habrán de acabar siendo digitales.

Los datos geográficos digitales tienen una serie de ventajas frente a los analógicos (además del mero hecho de que podemos incorporarlos a nuestro SIG), y suponen, como sucede en muchos otros campos, un salto cualitativo importante. Estas ventajas pueden resumirse en las siguientes:

- **Sencillez de actualización**. La cartografía digital es editable, y esto simplifica enormemente la introducción de cambios. Si en una capa con información catastral cambia la frontera de una parcela, basta modificar esta frontera. En un mapa analógico habría que rehacer todo el mapa y volver a imprimirlo.
  Además, y gracias a la división en capas, pueden actualizarse a distintos ritmos las distintas variables, pues son independientes y pueden modificarse por separado.
  Haciendo una analogía con el mundo editorial, piénsese en un diario impreso, con una única edición al día, en la que se ha de esperar al día siguiente para introducir todas las noticias que se vayan produciendo durante esa misma jornada. En su equivalente digital, la información se actualiza prácticamente en tiempo real, y podemos conocer las noticias mucho antes, pues es más sencillo actualizar esa página que volver a poner la imprenta en marcha.
  Es asimismo muy importante el hecho de que, gracias a los sistemas que centralizan el acceso a los datos, esta edición y actualización de datos pueden hacerla varias personas de modo concurrente. Esto no resulta posible en el caso de cartografía impresa, donde frecuentemente se encuentra el problema de que una cartografía de uso interno en una organización (por ejemplo, un ayuntamiento que guarda un inventario de su mobiliario urbano) ha sido editada por varias personas (el operario que sustituye un elemento de ese mobiliario luego lo registra en su inventario, y en un instante distinto otro operario puede añadir en su propio mapa la localización de un nuevo elemento añadido), siendo necesario después unir todas las modificaciones, lo cual no siempre resulta sencillo o incluso posible.
  Si varias personas trabajan con cartografía impresa de una zona, cada una de ellas tendrá su propio mapa. Con la cartografía digital, todos pueden obtener la cartografía de un repositorio central, de tal modo que si la editan, están editando una única versión, y no es necesario después poner en común todas sus aportaciones para crear una nueva cartografía actualizada.
- **Facilidad de distribución**. Resulta más sencillo y menos costoso distribuir cartografía digital que analógica, por ejemplo a través de Internet. Volviendo al ejemplo del diario, las noticias se actualizan y se ponen en la Web, de donde cada lector las descarga de inmediato. El diario impreso requiere una cadena de distribución más

costosa, desde la imprenta hasta el punto de venta.

- **Espacio de almacenamiento**. Se generan actualmente volúmenes ingentes de datos que además, y gracias a que son más fáciles de actualizar, se producen con una frecuencia mucho mayor. No obstante, un soporte digital puede almacenar una enorme cantidad de estos ocupando una fracción del espacio físico. En un ordenador dotado de una buena capacidad de almacenamiento caben los contenidos de una cartoteca y los de la hemeroteca de ese diario del que hablamos. Las mismas cartoteca y hemeroteca en formato impreso requieren edificios enteros.
- **Facilidad y precisión de análisis**. Como ya veremos en la parte correspondiente, el salto cualitativo que se da en el campo del análisis es enorme. Podemos hacer con los datos geográficos digitales cosas que no eran posibles con los analógicos y, mejor aún, podemos automatizar estos análisis. Asimismo, la precisión es mayor, ya que depende únicamente de los datos y la precisión intrínseca de estos, pero no de la operación de análisis (piénsese en un mapa impreso y una serie de operarios midiendo la longitud de un río sobre él. Es probable que lleguen a resultados similares pero no idénticos. Con cartografía digital, cualquier operario, y en cualquier SIG —suponiendo que implementan todos las mismas fórmulas— llegaría al mismo resultado exacto).
- **Facilidad de mantenimiento**. Aunque no se introduzcan modificaciones y no se actualicen los datos, el formato digital hace más fácil su conservación. La degradación del soporte no degrada directamente el dato en sí, haciéndole perder calidad. La degradación del soporte analógico (el papel), sí que lo hace. Además, los datos digitales pueden replicarse con suma facilidad, por lo que su persistencia está garantizada en mayor medida y a un menor coste que la de los datos analógicos.

Así pues, disponemos para nuestro trabajo en nuestro SIG de datos analógicos y datos digitales, siendo estos últimos los que necesitamos, y que presentan las ventajas anteriormente descritas frente a los primeros. En las siguientes secciones, veremos con detalle todos los distintos tipos de datos geográficos, tanto digitales como analógicos, la forma en que se obtienen, sus características, cómo se incorporan a un SIG, y en general todo aquello que resulte de interés para una mejor comprensión y uso de los mismos.

6.3. Fuentes primarias y fuentes secundarias

Como hemos visto, algunos datos que utilizamos en un SIG son de tipo analógico, mientras que otros son de tipo digital. En algunos casos (generalmente en los analógicos), estos datos no han sido tomados pensando en su utilización en un SIG, y nos van a servir de base para obtener otros que sí pueden emplearse directamente dentro de un SIG. Por el contrario, existen otros datos que ya han sido recogidos considerando su utilización dentro de un Sistema de Información Geográfica, y la forma en la que se presentan ya es adecuada para incorporarlos en este y trabajar con ellos.

En base a lo anterior, se define una forma distinta de clasificar los datos espaciales con los que trabajamos en un SIG: datos primarios (o procedentes de una fuente primaria) y datos secundarios (o procedentes de una fuente secundaria) [41].

Los datos primarios son aquellos que podemos emplear en un SIG y que, en su forma original, ya son susceptibles de ser sometidos a las operaciones de manejo y análisis que incorporan los SIG. En este grupo encontramos las imágenes digitales o los datos obtenidos con GPS, todos

ellos recogidos ya en origen de forma adecuada para su empleo directo en un SIG.

Por su parte, los datos secundarios derivan de algún otro tipo de dato previo, el cual no es adecuado para su empleo en un SIG. Entre estos incluimos las versiones digitales de los mapas clásicos (veremos en breve cómo se lleva a cabo esa conversión de un documento analógico a uno digital), así como los datos procedentes de un muestreo o levantamiento tradicional. Otros provenientes de cartografía impresa, tales como capas de elevaciones, también se incluyen en este grupo.

Al desarrollar las fuentes de datos en este capítulo, se tratarán tanto fuentes primarias como secundarias, y en el caso de estas últimas se tratarán a su vez las formas en las que a partir de estas pueden derivarse datos digitales que puedan ya ser incorporados a un SIG.

6.4. Teledetección

La primera fuente de datos que trataremos en este capítulo es la teledetección. Entendemos por teledetección el estudio y medida de las características de una serie de objetos (en nuestro caso elementos de la superficie terrestre) sin que exista contacto físico [42, 43, 44]. Para ello, se miden las perturbaciones que el objeto provoca en su entorno, principalmente las de tipo electromagnético.

Tradicionalmente, la teledetección se ha estudiado como una materia complementaria pero en cierto modo separada de los Sistemas de Información Geográfica. Ello es debido principalmente a que se trata de una materia muy extensa cuyo desarrollo se ha producido en cierta parte de forma ajena al de los SIG. No obstante, a medida que ambos campos se han ido desarrollando, la convergencia entre SIG y teledetección se ha ido haciendo cada vez más evidente. No solo las aplicaciones SIG incorporan elementos para el manejo, tratamiento y análisis de datos procedentes de la teledetección, sino que las formulaciones de ambos ámbitos contienen elementos similares.

La teledetección es hoy en día un elemento clave para la formación en SIG, y como tal debe incluirse en un libro como este. Los bloques tradicionales en los que se divide el temario fundamental de la teledetección no incorporan únicamente el registro de la información y la creación de los datos, sino también su proceso posterior, interpretación y tratamiento. Este último no se detalla en este capítulo, sino en la parte dedicada al análisis, integrado junto con otras formulaciones similares para proceso de imágenes.

Aunque el aprovechamiento de los productos actuales de la teledetección depende en gran medida del concurso de los SIG y sus capacidades de análisis, los procesos de teledetección aplicados al ámbito cartógráfico y el análisis espacial se remontan a tiempo atrás, concretamente a la mitad del siglo XIX. Fue entonces cuando se tomaron las primeras fotografías aéreas, uniendo el recién desarrollado campo de la fotografía con la utilización de globos aerostáticos como medio para situar el aparato fotográfico a una altura suficiente que permitiera obtener las imágenes.

Las fotografías aéreas fueron el primer producto de la teledetección, pero hoy en día existen otros que, basados en esa misma idea de registro de información, pueden ser empleados como fuentes de datos espaciales dentro de un SIG. Para comprenderlos, estudiemos algo más en detalle los elementos del proceso de teledetección, los cuales se representan de forma esquemática en la figura 6.1.

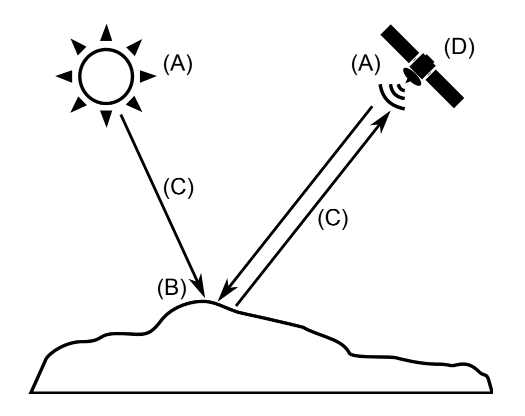


Figura 6.1: Esquema de un sistema de teledetección.

Estos elementos son los siguientes:

- **Una fuente de radiación (A)**. Puede ser de origen natural o artificial. La radiación emitida por dicha fuente llega al terreno y sufre una perturbación causada por los elementos de este, siendo esta perturbación el objeto de estudio de la teledetección. Los propios objetos pueden ser también emisores ellos mismos de radiación.
- **Unos objetos (B) que interaccionan con la radiación** o la emiten, según lo anterior.
- **Una atmósfera (C)** por la que se desplaza la radiación, tanto desde la fuente hasta el objeto como desde el objeto hasta el receptor. La atmósfera también interactúa con la radiación, introduciendo igualmente perturbaciones en ella.
- **Un receptor (D) que recoge la radiación** una vez esta ha sido perturbada o emitida por los objetos. El receptor va a generar como producto final una imagen (en términos de un SIG, una capa ráster), en cuyas celdas o píxeles se va a contener un valor que indica la intensidad de la radiación. Estos valores son valores enteros que indican el nivel de dicha radiación dentro de una escala definida (habitualmente valores entre 1 y 256), y se conocen dentro del ámbito de la teledetección como *Niveles Digitales*.

Para estudiar los dos primeros, estudiaremos los fundamentos físicos relativos a la radiación y a la la interacción entre esta y la materia, mientras que para el estudio del sistema receptor analizaremos dos de sus componentes por separado: sensores y plataformas.

La interacción de la atmósfera interesa de cara a eliminar su efecto, ya que lo que resulta de interés en general son los objetos en la superficie terrestre, no la atmósfera como tal. Eliminar esta influencia de la atmósfera es parte de los procesos posteriores que se realizan con la imagen y que incluyen también, como se mencionó anteriormente, la interpretación y otros procedimientos diversos sobre esta. Todos ellos se tratan, tal y como se dijo, en un capítulo independiente dentro de la parte de procesos.

6.4.1. Fundamentos físicos

Es necesario conocer los conceptos fundamentales sobre la radiación y su interacción con la materia (los objetos de la superficie terrestre) para poder entender cómo, utilizando la radiación de una fuente dada, se crea una imagen como resultado final en un proceso de teledetección.

La radiación electromagnética

La radiación electromagnética es una de las cuatro fuerzas fundamentales de la naturaleza[1] y deriva del campo electromagnético, el cual es ejercido por las partículas cargadas eléctricamente. Para explicar esta existen dos modelos conocidos como *modelo ondulatorio* y *modelo de partículas*. Según el primero, que será en el que profundicemos, la radiación electromagnética es producto de las alteraciones en los campos eléctrico y magnético, que generan dos ondas ortogonales entre sí, correspondientes a cada uno de los campos anteriores (Figura 6.2).

[1]Las otras tres son la gravitatoria, la interacción nuclear débil y la interacción nuclear fuerte

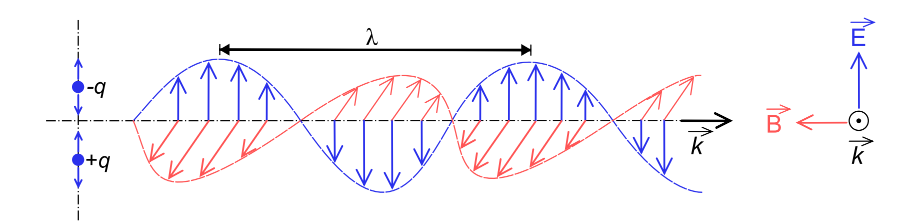


Figura 6.2: Ondas correspondientes a los campos magnético y eléctrico, ortogonales entre sí (Tomado de Wikipedia).

Estas ondas se desplazan a la velocidad de la luz, y se pueden describir con los parámetros habituales, tales como la longitud de onda o la frecuencia. Una mayor longitud de onda (y, por tanto una menor frecuencia) tiene asociada una mayor energía de la radiación.

La radiación electromagnética puede cubrir de forma continua todo un amplio rango de valores de longitudes de onda. Este rango se conoce como *espectro electromagnético*. Pese a la continuidad de sus valores, es habitual agruparlos en regiones, discretizando la amplitud del espectro, ya que las radiaciones en longitudes de onda cercanas presentan a su vez comportamientos similares en muchos sentidos. En la figura 6.3 se muestra un esquema del espectro electromágnético y sus principales regiones de interés.

Dentro de estas regiones, son de destacar las siguientes:

- **Rayos** γ <0.03 nm.
- **Rayos X** (0.03 nm - 3 nm).
- **Ultravioleta** (3 nm - 0.3 μ).
- **Visible** (0.3 μ - 0.7 μ). Se corresponde con las radiaciones que pueden ser detectadas por el ojo humano o por aparatos tales como una cámara fotográfica común. Observando la figura 6.3 puede apreciarse que esta región representa una porción muy pequeña del total del espectro. Existen muchas otras regiones que no podemos «ver» pero que, empleando la tecnología adecuada, sí que pueden aprovecharse para crear imágenes dentro de un proceso de teledetección, siendo de gran utilidad para el estudio de numerosos procesos.
Las distintas longitudes de onda dentro de esta región son las responsables de los distintos colores que percibimos. Así, por ejemplo, el azul se corresponde con el rango entre 0.4 μ y 0.5 μ, mientras que el verde lo hace con el rango entre 0.5 μ y 0.6 μ
- **Infrarrojo cercano** (0.7 μ - 1.3 μ).
- **Infrarrojo medio** (1.3 μ - 8 μ).
- **Infrarrojo lejano o térmico** (8 μ - 14 μ). Dentro de esta región se encuentran principalmente las radiaciones emitidas por los cuerpos debido a su temperatura.
- **Microondas** (1 mm - 25 cm).

En el capítulo 17 estudiaremos para qué tipo de análisis resulta útil cada una de las regiones del espectro, cuando veamos como analizar las imágenes procedentes de la teledetección.

Como ya se dijo en el capítulo 5, las imágenes como capas ráster presentan habitualmente la particularidad de tener varias bandas. En lugar de un único valor para cada celda, existen n valores, uno por cada banda. Esto es así porque la imagen recoge la intensidad de la radiación (recordemos que esto se denominaba Nivel Digital) dentro de una amplitud dada del espectro, y a su vez subdivide esta en distintas franjas. Los Niveles Digitales de cada banda corresponden a

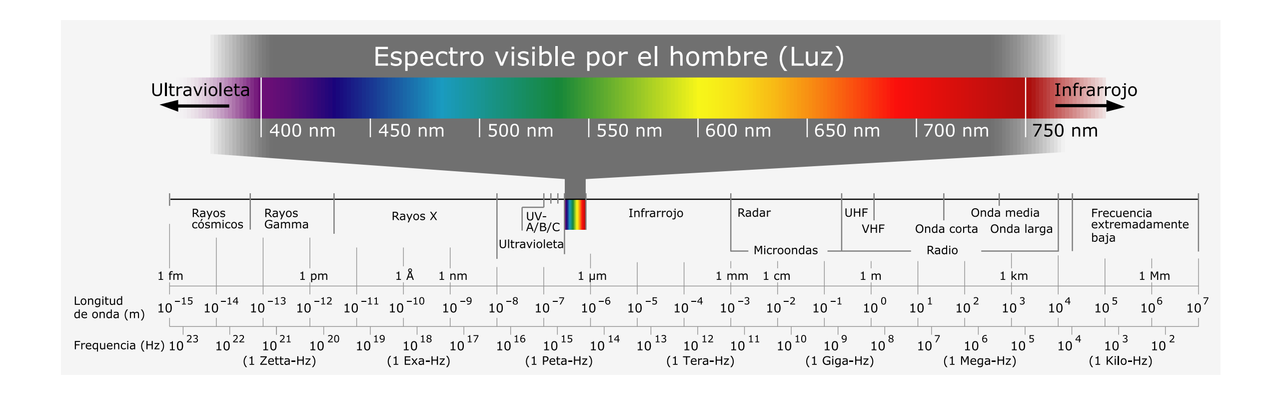


Figura 6.3: Espectro electromagnético y sus principales regiones de interés (Tomado de Wikipedia).

la intensidad dentro de una de esas franjas del espectro en particular.

Interacción entre radiación y materia

La radiación emitida por una fuente de radiación es alterada por la presencia de los distintos objetos, que interactúan con ella. Independientemente de su procedencia, para toda radiación se dan tres fenómenos fundamentales al alcanzar un objeto:

- **Absorción**. El objeto toma la energía de la radiación.
- **Transmisión**. La radiación atraviesa el objeto y continua su camino.
- **Reflexión**. La radiación «rebota» en el objeto y vuelve al espacio.

Estos tres fenómenos se dan en diferente proporción en función de las características del objeto y de la radiación. Para una longitud de onda dada, existe, pues, un porcentaje de la radiación que es absorbida por el objeto, otra que se transmite a través de él y otra que es reflejada. La parte que interesa a efectos de la teledetección es aquella que se refleja en el objeto, ya que esta es la que posteriormente puede recogerse y emplearse para la generación de las imágenes.

La proporción en la que los tres procesos anteriores se dan en un objeto no es la misma para todas las radiaciones. Un objeto puede absorber una gran parte de la radiación dentro de una región del espectro y sin embargo reflejar la mayoría de ella en una región distinta. Es por ello que, en función del análisis que se desee realizar, debe trabajarse con imágenes que traten una u otra región.

Igualmente, una imagen con varias bandas contiene información sobre la intensidad de la radiación reflejada en distintas partes del espectro. Puesto que cada objeto refleja de forma diferente la radiación en cada una de esas partes, pueden igualmente emplearse para identificar objetos particulares si se conoce la respuesta de estos en determinadas bandas. Por ejemplo, si sabemos que los objetos que buscamos reflejan gran cantidad de radiación en todas las longitudes de onda excepto en un rango concreto. Aparece así el concepto de *firma espectral* como la respuesta característica de un tipo de objeto dentro del espectro electromagnético. Veremos mucho más al respecto en el capítulo 17, así como en el 21, donde estudiaremos una aplicación habitual de dichas firmas espectrales.

Además de la interacción con los objetos que se pretenden estudiar, la radiación interactúa con la atmósfera. Esta interacción afecta al resultado y es una variable a considerar en ciertas operaciones posteriores con las imágenes. Veremos más sobre la interacción entre radiación y atmósfera en el apartado 17.2, cuando tratemos esas operaciones.

6.4.2. Sensores y plataformas

En un sistema de teledetección, dos son los elementos tecnológicos principales que lo definen: el *sensor* y la *plataforma*.

El sensor es el elemento que incorpora la capacidad de «leer» la radiación electromagnética y registrar su intensidad dentro de la una zona concreta del espectro. En palabras más sencillas, es el aparato que nos permite «tomar» la imagen, y puede ir desde una simple cámara fotográfica hasta un sensor más especializado capaz de tomar cientos de bandas en una región del espectro de gran amplitud.

La plataforma, por su parte, es el medio en el que se sitúa el sensor y desde el cual se realiza la observación. Los dos tipos principales de plataformas son aquellas situadas dentro de la atmósfera terrestre (aviones en su mayoría, aunque también en otros medios tales como globos aerostáticos) y aquellas situadas fuera de la atmósfera (a bordo de satélites).

Las características de estos dos elementos definen las del sistema en su conjunto, así como las propiedades de sus productos derivados y la utilidad que estos presentan.

Plataformas

La plataforma es el medio en el que se transporta el sensor, y condiciona las mediciones efectuadas por este, ya que establece la distancia a la que el sensor se sitúa del elemento registrado (la superficie terrestre). Esta distancia puede ser del orden de algunos centenares de metros o unos pocos kilómetros, o bien de muchos kilómetros. En el primer caso, la plataforma más habitual es el avión, mientras que en el segundo caso lo más frecuente es el uso de satélites.

Los aviones son las plataformas clásicas a bordo de las cuales se montaban originariamente las cámaras empleadas para la realización de fotografías aéreas. Hoy en día pueden montarse igualmente otros sensores más complejos y modernos a bordo de aeronaves.

Las ventajas del empleo de aviones como plataformas de teledetección son las relacionadas con la disponibilidad de la plataforma, que es mucho mayor que en el caso de emplear satélites. Podemos (dentro de lo razonable) escoger cómo, cuándo y dónde efectuar un vuelo y tomar imágenes, mientras que en caso de satélites la disponibilidad viene condicionada por numerosos factores y es muy reducida.

Respecto a los inconvenientes, pueden citarse entre ellos la inestabilidad de la plataforma y la dependencia de las condiciones del clima, que pueden afectar a la propia estabilidad y a la calidad de los resultados, o incluso impedir la realización del vuelo. Por ser plataformas de baja altura, no pueden abarcar superficies tan amplias como los satélites, requiriendo más tiempo para cubrir una zona dada.

Por su parte, los satélites artificiales presentan unas características distintas como plataformas de teledetección, siendo muy útiles para la teledetección sobre la superficie terrestre. Es habitual que a bordo de un mismo satélite coexistan diversos sensores, de forma que una única plataforma transporta varios de ellos.

A diferencia de un avión, un satélite no puede dirigirse a voluntad (no puede pilotarse), y su movimiento es una característica inherente que viene definida por una serie de parámetros. Estos

parámetros se conocen como *parámetros orbitales*, pues definen la órbita descrita por el satélite en torno a la Tierra.

Las órbitas pueden clasificarse en función de su eje de rotación en tres tipos:

- **Ecuatoriales**, si se sitúan en el mismo plano en el ecuador terrestre.
- **Polares**, si se sitúan en un plano que contiene al eje de rotación terrestre.
- **Semipolares**, si la órbita es oblicua al eje de rotación

Con un criterio distinto, atendiendo a la forma en que se produce el movimiento, distinguimos dos tipos de órbitas para un satélite:

- **Geosíncronas**. El satélite se sitúa sobre un punto fijo de la Tierra y su movimiento sigue al de rotación de esta. Es decir, no existe movimiento relativo entre dicho punto de la superficie terrestre y el satélite. Todas las imágenes que se toman desde el satélite tendrán así el mismo encuadre y cubrirán una extensión idéntica. La altura del satélite es fija, siendo esta de 35.786 km, ya que esta altura hace que la velocidad del satélite se corresponda con la de rotación de la Tierra.
  La ventaja de este tipo de satélites es que, por situarse siempre sobre un punto y siempre teniendo visión sobre una zona dada, se pueden actualizar con mucha frecuencia las imágenes. El inconveniente principal radica en el hecho de que las zonas alejadas del punto sobre el que se sitúa el satélite tendrán mala cobertura, y existirán zonas no cubiertas de las que no resultará posible obtener imágenes con los sensores montados a bordo de dicho satélite. Pese a que un sensor sobre un satélite con órbita geosíncrona cubrirá una gran porción de la superficie terrestre (debido a la elevada altura a la que ha de situarse para tener dicha órbita), no resulta posible, como es lógico, cubrir toda ella y hacerlo además en las mismas condiciones en todas las zonas.
- **Heliosíncronas**. Las órbitas heliosíncronas son generalmente polares. Mientras el satélite recorre la órbita, la Tierra efectúa su movimiento de rotación, lo cual hace que a cada vuelta de la órbita se cubran zonas distintas. De esta forma, se consigue dividir la totalidad de la superficie terrestre en bandas que se van recorriendo sucesivamente hasta que el satélite vuelve a situarse en el mismo punto inicial. Las órbitas están diseñadas de tal manera que ese regreso al punto inicial se produce a la misma hora solar exacta que en el anterior ciclo, de forma que las imágenes tomadas en un punto dado son registradas siempre a la misma hora y en condiciones similares de iluminación. Para que sea posible realizar una órbita de este tipo, el satélite debe situarse entre 300 y 1500 km de altura.
  La figura 6.4 muestra un ejemplo de la forma en que un satélite con una órbita heliosíncrona barre toda la superficie de la Tierra.
  Debido al movimiento que causa las distintas franjas, los satélites con este tipo de órbitas pueden cubrir toda la superficie terrestre, algo que no es posible con los de órbita geosíncrona. No obstante, una vez que se toma una imagen de una zona, la plataforma no regresa a ella hasta que se concluye todo el ciclo, habiendo transcurrido un periodo de tiempo que se conoce como *periodo o intervalo de revisita*. Debido a ello, la actualización de las imágenes no puede ser tan inmediata como en el caso de satélites geosíncronos.

Sensores

Montado a bordo de cualquiera de los tipos de plataformas que hemos visto en el apartado anterior, el sensor es el encargado de registrar la radiación electromágnetica procedente de la zona estudiada y «tomar» la imagen.

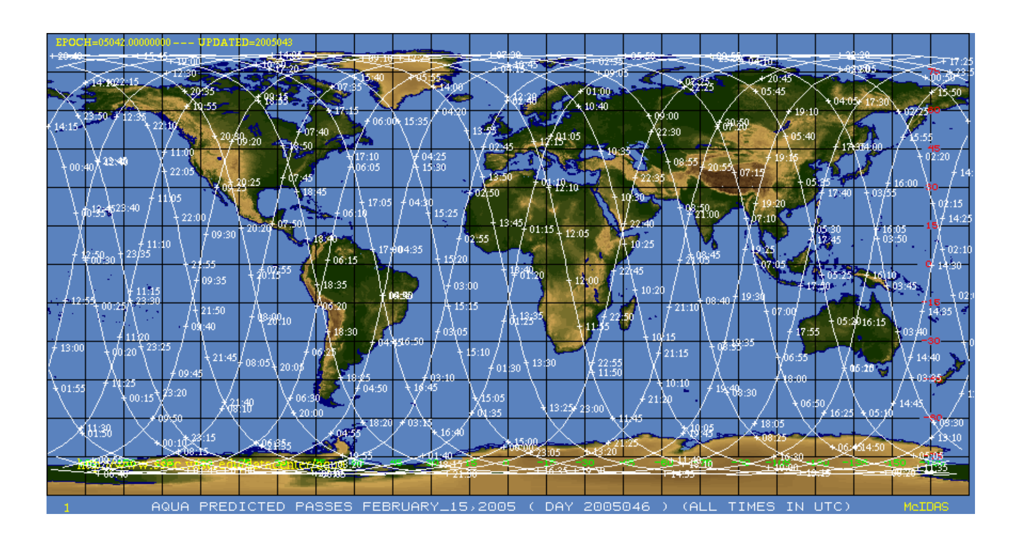


Figura 6.4: Esquema de barrido de un satélite con órbita heliosíncrona

Existen diversas formas de clasificar los sensores. Una división básica es la que distingue sensores *activos* y sensores *pasivos*. Como ya sabemos, la radiación que recoge el sensor es el resultado de una fuente de radiación electromagnética, cuyas emisiones interactúan con el medio, que refleja una parte de las radiaciones que le llegan. Los sensores pasivos aprovechan las fuentes de radiación existentes en la naturaleza (fundamentalmente el Sol) y se limitan a recoger la radiación de dichas fuentes reflejada por los elementos del medio, o la que estos elementos emiten por sí mismos. El sensor no produce ningún tipo de radiación de por sí. Por el contrario, los sensores activos sí emiten radiación, y recogen dicha radiación tras ser reflejada por los elementos del medio.

La diferencia fundamental entre estos dos tipos de sensores es que los activos pueden funcionar en cualquier instante y no dependen de la condiciones atmosféricas o el momento del día. De la misma forma que no podemos tomar una fotografía de noche sin luz, y no podemos ver el suelo desde un avión cuando hay nubes, no podemos utilizar un sensor pasivo en esas condiciones para tomar una imagen. Sin embargo, sí podemos hacer una fotografía de noche si disponemos de un flash, ya que la propia cámara emite la luz que necesita. La filosofía de un sensor activo es en cierta medida similar al caso de la cámara con flash.

Los sensores activos emiten su propia radiación, por lo que no es necesario que existan fuentes externas (no es necesaria la luz solar). Asimismo, los elementos atmosféricos tales como las nubes, que afectan a la radiación visible, no afectan a otros tipos de radiación, permitiéndoles una operatividad total en la gran mayoría de condiciones. Por ello, los sensores activos suelen trabajar en el rango de microondas (frente a los sensores pasivos, que lo hacen principalmente en las regiones del visible y el infrarrojo), ya que estas son capaces de atravesar la atmósfera en prácticamente todas las condiciones, presentando así ventajas frente a los sensores pasivos en este aspecto.

Aunque el producto habitual de la teledetección son las imágenes, entendidas estas como

algo *visual*, algunos sensores no forman tales imágenes, y los valores que recogen no son las intensidades de la radiación reflejada por el terreno en una longitud de onda dada. Es decir, no se corresponderían con el concepto de Nivel Digital ya presentado. Este tipo de resultados son habituales en los sensores de tipo activo, en los que la radiación que el propio sensor emite es recogida tras reflejarse en el terreno, pero la variable que se mide de ella no es su intensidad sino, por ejemplo, el tiempo que tarda en regresar. Planteamientos como estos permiten la generación de capas de datos que no son imágenes como tales, como es el caso de las capas de elevación (Modelos Digitales de Elevaciones), ya que el tiempo de retorno está directamente relacionado con la distancia recorrida por la radiación, y este con el relieve del terreno.

Estos sensores, no obstante, operan de un modo similar a lo que ya conocemos, y se consideran igualmente dentro del ámbito de la teledetección, pues se adscriben a la definición de esta dada al principio de este apartado. Veremos igualmente ejemplos de algunos de ellos cuando veamos más adelante algunos sensores de particular relevancia, ya que tienen una gran importancia en la actualidad para la generación de cartografía, como por ejemplo la ya citada de elevaciones.

El radar [2] es la tecnología más importante dentro de este grupo. El sensor envía pulsos de radio, y posteriormente recoge estos midiendo su intensidad y pudiendo calcular también la distancia al objeto.

Una técnica más moderna pero similar al radar es el denominado LiDAR [3], que emplea pulsos de láser. El LiDAR es en la actualidad la tecnología más avanzada para la creación de cartografía de elevaciones, y dentro de este campo ha supuesto una verdadera revolución, ya que obtiene resoluciones muy elevadas, tanto horizontales como verticales (resolución en los valores de elevación calculados).

Los sistemas modernos de LiDAR son capaces de proporcionar además varios retornos, de modo que, si el sensor sobrevuela una zona arbolada, se tiene información sobre la distancia a la copa y la distancia al suelo, ya que parte del láser atraviesa la copa y alcanza el terreno. Este tipo de resultados supone un salto cualitativo con respecto a los obtenidos con otras tecnologías. Esto permite no solo estudiar el terreno, sino derivar otros parámetros tales como la altura de la vegetación [45]. Asimismo, debido a su precisión, permite recoger elementos del terreno que con otros sistemas no resulta posible registrar, tales como edificios. A modo de ejemplo, la figura 6.5 muestra un modelo del World Trade Center el 27 de septiembre de 2001, creado a partir de datos LiDAR.

En la terminología del LiDAR, la imagen correspondiente al primer retorno (el de los puntos más altos) se conoce como Modelo Digital de Superficie (MDS), mientras que el correspondiente a la altura del suelo se conoce como Modelo Digital de Elevaciones (MDE). Veremos mucho acerca de los MDE en posteriores capítulos de este libro.

En [46] puede encontrarse una buena descripción del proceso de creación de capas de elevación a partir de datos LiDAR.

Además de la división entre activos y pasivos, otra forma de clasificar los sensores es en función de la forma en la que registran la imagen.

Algunos sensores poseen un único detector de radiación que no cubre todo el ancho de la franja del terreno que se pretende recoger. Por medio de espejos oscilantes, se envía a este detector la radiación procedente de los distintos puntos a lo ancho de esa franja, de forma que se

[2] Acrónimo de *Radio Detection and Ranging*, detección y medición a partir de ondas de radio
[3] Acrónimo de *Light Detection and Ranging*, detección y medición de distancias a partir de luz

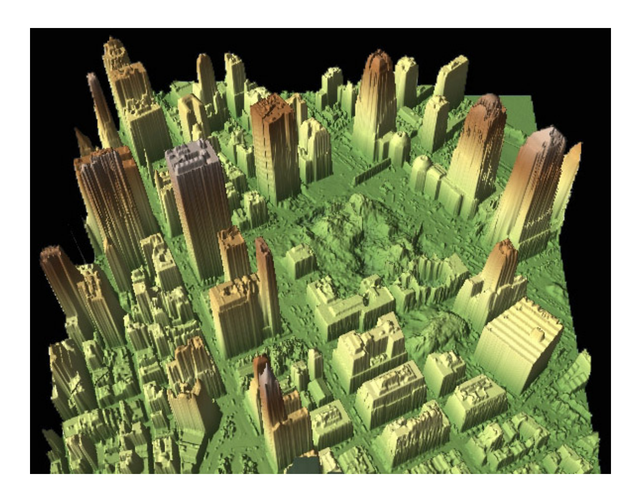


Figura 6.5: Modelo del World Trade Center realizado a partir de datos LiDAR tomados el día 27 de septiembre de 2001 (Fuente: NOAA/U.S. Army JPSD)

van recogiendo los píxeles de la imagen uno a uno, recorriendo esta de un lado a otro (Figura 6.6a). Estos sensores se denominan *de barrido*.

Los denominados sensores *de empuje* (Figura 6.6b) eliminan la necesidad de utilizar espejos móviles, ya que poseen un número mayor de detectores que permiten cubrir todo el ancho de la imagen. Por ello, esta se va registrando no píxel a píxel, sino línea a línea.

Resoluciones

Uno de los parámetros principales que definen las propiedades de un sistema de teledetección son las *resoluciones*. Estas establecen el nivel de detalle de los productos que el sistema genera, determinando este en las distintas magnitudes en las que el sistema opera. Las resoluciones dependen del sensor y de la plataforma como binomio operativo, y de las características propias de ambos. Distinguimos cuatro resoluciones, a saber:

- **Resolución espacial**. Indica la dimensión del objeto más pequeño que puede distinguirse en la imagen. En líneas generales es el equivalente al tamaño de píxel[4] es decir, a la dimensión real que un píxel de la imagen tiene sobre el terreno. La resolución espacial está en función de la capacidad resolutiva del sensor y las características de la plataforma tales como la altura a la que se sitúa. Asimismo, la resolución espacial esta relacionada con la superficie que cada imagen cubre sobre el terreno. El concepto de *Campo Instantáneo de Visión*[5] indica el ángulo de visión que abarca el sensor, y se utiliza habitualmente en este sentido. El *Campo Instantáneo de Visión en Tierra*[6] expresa esta misma idea pero en unidades de longitud sobre el terreno, y es función del IFOV y la altura a la que se encuentre el sensor.

[4]Desde un punto de vista formal, no ha de ser necesariamente así, ya que la imagen puede tomarse originalmente con unas características y después, mediante operaciones matemáticas (veremos estas en el capítulo 15), modificar el tamaño de píxel. Aunque este tamaño sea menor al original, los objetos de menor dimensión que podrán discernirse en esa imagen no serán iguales a ese tamaño, sino mayores.

[5]Instantaneous Field of View (IFOV)

[6]Ground Instantaneous Field of Vision (GIFOV)

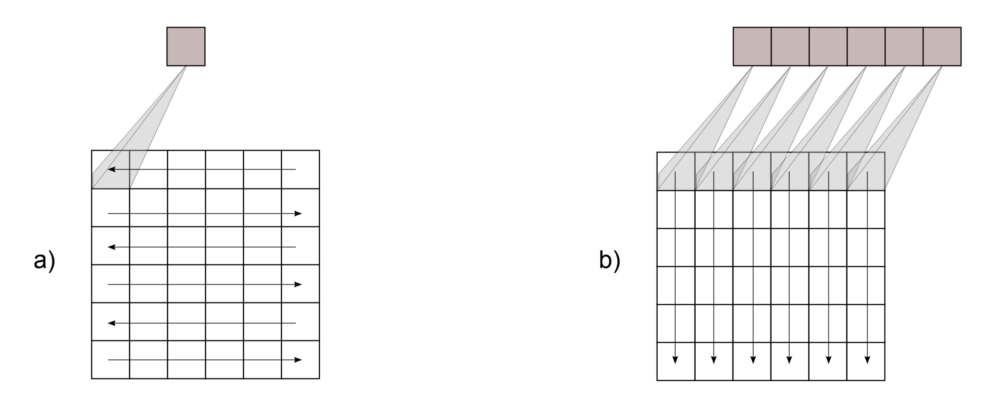


Figura 6.6: Esquema de funcionamiento de un sensor de barrido (a) y uno de empuje (b)

En el diseño de la órbita de un satélite debe tenerse en cuenta el campo de visión del sensor para optimizar el ciclo de toma de imágenes, así como para evitar que las distintas franjas que este cubre queden sin solaparse y existan zonas de las que no se tomen imágenes.

- **Resolución espectral**. Todo sensor cubre una región particular del espectro y almacena esta mediante un número dado de bandas. La región del espectro abarcada y el número de bandas son los elementos que definen la resolución espectral. Esta será elevada si el número de bandas es alto, ya que cada banda cubrirá un rango de frecuencias de menor amplitud. De este modo, la información de dos frecuencias cercanas puede separarse, ya que estas serán recogidas en bandas distintas, mientras que si el número de bandas es menor pertenecerán a la misma banda y no podrá hacerse distinción alguna (la resolución será menor).

  En función del número de bandas, pueden clasificarse las imágenes y los sensores que las generan. Una imagen en blanco y negro contiene una única banda. Las imágenes en color contienen tres bandas, correspondientes a las frecuencias del rojo, el verde y el azul. Existen igualmente sensores con algunas bandas adicionales como la del infrarrojo, que en total generan un número de bandas no superior a diez. Todas estas imágenes se conocen como *multiespectrales*.

  Las imágenes *superespectrales* tienen una mayor resolución espectral (bandas más estrechas), y cubren una zona del espectro más amplia, no limitándose al rango visible o el situado inmediatamente junto a este. Por ello, su número de bandas es mayor, generando imágenes con varias decenas de ellas.

  Por último, las imágenes *hiperespectrales* presentan más de cien bandas, lo cual permite una caracterización espectral sumamente precisa.

- **Resolución radiométrica**. Para cada una de las bandas que produce un sensor (asociada esta a una determinada región del espectro según su resolución espectral), el dato recogido, que constituye su Nivel Digital, indica la intensidad correspondiente a esa región. El nivel de detalle con el que puede medirse esa intensidad es el que define la resolución radiométrica del sensor. El número de Niveles Digitales distintos que pueden recogerse es la medida de la resolución espacial, y habitualmente es una potencia de dos (de la forma 2^n). Tanto las imágenes en blanco y negro como las imágenes en color trabajan con 256 (2^8) niveles, ya que este es el valor más cercano al número de diferentes intensidades que el

ojo humano puede diferenciar[7]. No obstante, los sensores de teledetección pueden tener una mayor resolución radiométrica (hasta 1024 o 2048 niveles), que si bien no se aprecia en la representación visual, sí que supone una diferencia en el tratamiento analítico de esos Niveles Digitales.

- **Resolución temporal**. Indica el tiempo que tarda el sensor en volver a tomar una imagen de una misma zona. Tiene sentido en el caso de sensores orbitales, que funcionan por ciclos, y que tras concluir este vuelven a comenzar la toma de imágenes en el mismo punto. En cada ciclo, el sensor cubre toda la superficie terrestre «barriendo» esta en franjas sucesivas.
  La resolución temporal depende de la altura a la que se encuentra la plataforma que monta el sensor, así como de la resolución espacial. Si el tamaño de las imágenes es reducido (GIFOV pequeño), las franjas son más estrechas y se requieren más para cubrir toda la superficie y volver a comenzar el ciclo, con lo que la resolución temporal será menor.

Parece lógico pensar que lo ideal en toda circunstancia sería disponer de imágenes procedentes de sistemas con altas resoluciones en cualquiera de las clases anteriores. De esta forma, tendríamos imágenes con gran detalle espacial, espectral y radiométrico, y actualizadas frecuentemente. No obstante, la tecnología actual no dispone de elementos que ofrezcan resoluciones elevadas en todas las magnitudes del proceso, y en la creación de los sensores se favorecen unas en detrimento de otras. Algunas resoluciones presentan además un cierto antagonismo, como hemos visto para las resoluciones espacial y temporal, con lo que no resulta viable que ambas sean elevadas simultáneamente.

Así, existen sensores con, por ejemplo, gran resolución espacial, en los cuales la resolución espectral no es tan elevada. Por el contrario, los sensores con mayor resolución espectral no suelen ofrecer un nivel de detalle espacial tan elevado como los anteriores. En ocasiones, una misma plataforma puede montar a bordo varios sensores, de tal forma que el conjunto de ellos ofrezca información detallada de forma global, pero un único sensor no proporciona resolución elevada en todas las variables.

Otro tipo de circunstancias relativas al sensor afectan igualmente a las resoluciones. Por ejemplo, aquellos sensores que trabajan con radiaciones de poca energía (en la región de las microondas) y son de tipo pasivo requieren una amplia extensión para recoger la suficiente energía como para poder ser detectada por dicho sensor. Por esta razón, su resolución espacial suele ser baja.

A la hora de utilizar imágenes de teledetección, debe considerarse qué tipo de resolución resulta de mayor interés para el proyecto que se lleva a cabo, teniendo en cuenta la escala de trabajo o el objetivo final que se persigue, entre otros factores. En base a esto, se escogerá uno u otro producto, que será el que ofrezca los valores de resolución más adecuados en conjunto.

Si se pretende localizar elementos de pequeño tamaño, es imprescindible trabajar con altas resoluciones espaciales. Si lo que se desea es clasificar una serie de zonas en función de sus características, la resolución espectral debe ser alta, ya que, como veremos, se usa la información de todas las bandas para dar esa clasificación, y un número mayor de bandas dará como resultado una mayor precisión.

[7] En el ámbito del tratamiento de imágenes esto se conoce como *profundidad de color*. Una mayor profundidad de color indica mayor número de colores posibles. Una pantalla normal de ordenador puede mostrar un total de 16.7 millones de colores distintos, que corresponden a las combinaciones entre los 256 posibles niveles de cada una de las tres bandas ($256^3 = 16,777,216$)

De igual modo, la detección de cambios de intensidad en una banda hace necesario que se trabaje con una buena resolución radiométrica, pero si lo que se desea es estudiar esos cambios a lo largo de un periodo corto de tiempo, trabajar con un sensor con gran resolución temporal se hace imprescindible.

En cada caso, las circunstancias particulares del trabajo condicionan la elección de uno u otro sensor, puesto que, como se ha dicho, un único sensor no ofrece elevadas resoluciones en todas las variables.

La utilización simultánea de datos de varios sensores en un proyecto es una alternativa en ciertos casos. Como veremos, existen técnicas que permiten combinar imágenes con alta resolución espacial e imágenes con alta resolución espectral, con objeto de obtener nuevas imágenes que combinen lo mejor de ambas y ofrezcan un nivel de detalle conjunto mayor. Estas técnicas realizan el proceso conocido como *fusión de imágenes*, el cual trataremos en el apartado 17.3.3, más adelante en este libro.

Además de lo anterior, un único sensor montado a bordo de un satélite puede operar en varios *modos* distintos. Es habitual que un sensor multibanda pueda registrar también imágenes de una sola banda, recogiendo en ella la intensidad de la radiación correspondiente a todo el espectro visible, de tal forma que genere una representación visual real. Estas se suelen representar habitualmente en escala de grises, resultando una imagen en blanco y negro.

Las imágenes de este tipo se conocen como *pancromáticas*[8], y suelen tener mayor resolución espacial, por lo que pueden emplearse para la fusión de imágenes señalada anteriormente. Así, un mismo sensor provee todos los datos necesarios para llevar a cabo ese proceso, tanto la imagen de gran resolución espacial (la pancromática) como la de gran resolución espectral (la imagen multibanda).

6.4.3. Principales sensores y productos

El número de diferentes productos provenientes de la teledetección es muy elevado en la actualidad. Ahora que ya conocemos los fundamentos del proceso y las principales características de un sistema de teledetección, es interesante mostrar un pequeño resumen de los principales productos disponibles.

A continuación se relacionan algunos de los sistemas de teledetección principales y las características de sus productos.

- **LANDSAT**. Se trata de un programa completo de adquisición de datos mediante teledetección, que ha lanzado hasta la fecha un total de siete satélites entre 1972 y 1999. Por ello, el volumen de datos recogido es enorme, y lo convierte en una de las fuentes de datos más ricas de entre las existentes en la actualidad.
  El último satélite, LANDSAT 7, tiene una órbita heliosíncrona y una resolución temporal de 16 días. A bordo de él se monta el sensor ETM+[9], que permite la obtención de imágenes pancromáticas con resolución de 15 metros, e imagenes multibanda con resolución de 60 metros. El sensor recoge un total de 8 bandas, y el tamaño de la imagen es de 170 × 183 km.

[8]El término *pancromático* deriva de la fotografía clásica, conociéndose así al tipo de película sensible a todas las longitudes de onda del visible. Por similitud de conceptos, se emplea el término también para hacer referencia a las imágenes digitales monobanda generadas por sensores según lo comentado anteriormente

[9]Enhanced Thematic Mapper Plus

[10]Thematic Mapper

[11]Multispectral Scanner

Los sensores TM[10] y MSS [11] se montan a bordo del satélite LANDSAT 5, todavía en funcionamiento y con una resolución temporal de 16 días. El sensor TM ofrece imágenes multibanda de 7 bandas con resolución de 30 metros, excepto en la banda del infrarrojo térmico, donde la resolución es de 120 metros. Las imágenes tienen un tamaño de 185 × 172 km.

- **IKONOS**. Este satélite, lanzado en 1999, monta un sensor con resolución de 1 metro para imágenes pancromáticas y 4 metros para imágenes multibanda (4 bandas). Las imágenes cubren una área de 11 × 11 km y el satélite tiene una resolución temporal de entre 3 y 5 días.
- **SPOT**[12]. Un conjunto de satélites lanzados inicialmente por la agencia espacial francesa, con especial énfasis en la recogida de información relativa a variables ambientales. De los cinco puestos en órbita, dos siguen actualmente en funcionamiento. El último de ellos, lanzado en 2002, monta el sensor HRG con capacidad de producir imágenes pancromáticas con resolución entre 2,5 y 5 metros, e imágenes multibanda con resolución de 10 metros. El periodo de revisita es de entre 1 y 4 días. Es de destacar que el sensor permite inclinaciones de hasta 27°respecto al nadir hacia ambos lados, por lo que puede cubrir una banda más ancha y tomar imágenes fuera del área determinada en cada instante por la órbita.
- **QuickBird**. Ofrece imágenes en pancromático y multibanda (azul, verde, rojo e infrarrojo cercano). Las primeras tiene una resolución de 60 cm y las multibanda de 2,4 metros, aunque combinando las dos ofrece imágenes en color con 60 cm de resolución. La órbita del satélite es heliosíncrona y la resolución temporal varía entre los 3 y 7 días. Cada imagen cubre una superficie de 16,5 × 16,5 km.
- **Aqua y Terra**. Dos satélites lanzados por la NASA dentro de un proyecto de ámbito internacional para la observación de la Tierra. Cada uno de ellos monta una serie de diversos sensores, que recogen información relativa al ciclo hidrológico (en el caso del Aqua) y la superficie terrestre (en el caso del Terra). Entre estos sensores cabe destacar el MODIS, a bordo de ambos, o el ASTER, a bordo del satélite Terra. ASTER [13] recoge información en 14 bandas distintas, con una resolución entre 15 y 90 metros, mientras que MODIS[14] es un satélite de menor resolución espacial (250, 500 o 1000 metros según la banda), 36 bandas y una resolucion temporal de 1 a 2 días.

  Además de los datos directos de los sensores, se proporcionan de forma gratuita numerosos productos derivados, lo que lo convierte en una fuente de datos de primer orden para un gran número de aplicaciones, especialmente las relacionadas con el estudio del medio, la vegetación, etc.
- **NOAA–AVHRR**[15]. Se encuentra principalmente enfocado al estudio de los océanos, aunque sus datos pueden aplicarse en muchos más estudios. El sensor tiene una resolución de 1,1 km, y proporciona imágenes de 5 bandas en las regiones del infrarrojo y el visible. La resolución temporal es de medio día, produciendo una imagen nocturna y otra diurna.
- **RADARSAT**. Desarrollado por la Agencia Espacial Canadiense, monta un radar de apertura sintética (SAR), y su principal propósito es el control de las variaciones ambientales y de los recursos naturales.

[12]Satellite Pour l' Observation de la Terre
[13]Advanced Spaceborne Thermal Emission and Reflection Radiometer
[14]Moderate Resolution Imaging Spectroradiometer
[15]Advanced Very High Resolution Radiometer

- **ERS–1 y ERS–2**. Desarrollados por la Agencia Espacial Europea. Al igual que el anterior, ambos están pensados para la observación medioambiental, y montan tanto sensores activos como pasivos.
- **SRTM**. La misión SRTM[16] es un proyecto internacional de gran envergadura destinado a la creación de una cobertura de elevaciones a nivel mundial. Recogiendo el relieve de todas las zonas situadas entre los 56 grados sur y los 60 grados norte de latitud. La resolución es de un segundo de arco (aproximadamente 30 metros) para Estados Unidos, y de unos 90 metros en el resto de zonas.
- **PlanetScope y SkySat**. Dos constelaciones de satélites desarrolladas y gestionadas por la empresa Planet Labs. PlanetScope está compuesta por más de 150 de los denominados *Doves*. Estos pequeños satélites —cuyo tamaño es similar al de una caja de zapatos—, con una resolución espacial media de 3.7 metros y productos multibanda (RGB y NIR), se dedican al continuo monitoreo de nuestro planeta. SkySat, por su parte, es una constelación de 16 satélites de alta resolución espacial cercana a los 50 cm, y un rápido periodo de revisita (menor de 24 horas). Los objetivos a fotografiar de los satélites SkySat son definidos por los propios clientes a través de un sistema de interfaz y API.

6.5. Cartografía impresa. Digitalización

La primera fuente de cartografía de la que se disponía en las etapas iniciales de los SIG era la cartografía impresa. No se trataba de elementos creados pensando en su utilización dentro de un SIG y su estructura no es la más adecuada para su uso como datos de trabajo en un SIG. Se trata, por tanto, de una clara fuente secundaria de datos espaciales. Aun así, su uso ha sido desde esos tiempos una constante dentro del ámbito SIG.

A pesar de que hoy en día disponemos de otras fuentes cartográficas, la cartografía impresa sigue siendo básica para trabajar con un SIG, ya que existe mucha información que solo se encuentra en este formato. De una u otra forma, es probable que un proyecto SIG implique en algún punto de su desarrollo la necesidad de recurrir a cartografía impresa y tratar esta para su inclusión dentro de un SIG.

Cuando hablamos de cartografía impresa, no hay que pensar únicamente en mapas o planos, sino también en imágenes tales como fotografías aéreas, las cuales, dependiendo de su antigüedad, pueden encontrarse disponibles tan solo en papel. Mientras que resulta posible adquirir estas en formato digital cuando se trata de fotografías más actuales, la tomadas por métodos analógicos correspondientes a vuelos más antiguos solo pueden adquirirse por regla general como un producto impreso.

Los procesos que permiten obtener un producto digital a partir de esas imágenes son costosos en tiempo y dinero, y es por ello que no todos los proveedores de estas ofrecen la posibilidad de adquisición de un producto digital. En esta sección veremos esos procesos, tanto si partimos de un mapa o plano como si partimos de una imagen o cualquier otro documento impreso que pueda contener información cartográfica, susceptible de ser convertida en una o varias capas.

Tanto mapas como fotografías aéreas pueden servir como fuente de información para crear o bien capas ráster o bien capas vectoriales, ya que la información que contienen puede de igual modo representarse según uno u otro modelo.

Un mapa o plano sobre un soporte impreso, sin embargo, dista considerablemente de ese

[16] Shuttle Radar Topography Mission

concepto de capa con el que trabajamos en un SIG. Suele contener información sobre distintas variables, tales como carreteras, elevación, núcleos urbanos, uso de suelo, y todas ellas en un único elemento cartográfico. Esas variables, que en un SIG manejaríamos como capas independientes, se presentan como un conjunto que será conveniente disgregar en base a esas distintas variables.

La creación de un dato digital a partir de un dato analógico como un mapa impreso se conocen como *digitalización*, y esta puede ser automatica o manual.

En la digitalización automática, el sistema (informático o mecánico) se encarga de generar los elementos digitales que ya podremos incorporar a un SIG, ahorrando trabajo al operador al automatizar la tarea. Este tipo de digitalización es muy habitual para el caso de obtener un resultado ráster mediante el proceso de *escaneo*. También resulta posible automatizar la digitalización para el caso vectorial, aunque requiere cierta labor por parte del operario y no es un proceso tan sencillo, pudiendo obtenerse resultados desiguales.

La digitalización manual requiere por parte del operario una definición explícita de los elementos a crear, y es por ello únicamente adecuada para obtener un resultado vectorial, trazándose las entidades (sean estas puntos, líneas o polígonos) manualmente mediante algún sistema que permita esa introducción de datos.

La elección de uno u otro tipo de digitalización no depende solo del tipo de capa que se desee obtener. Tanto la digitalización manual como la automática, tienen cada una de ellas su propias ventajas. En el caso ráster la opción manual no es viable, pero al digitalizar un mapa para obtener una capa vectorial puede ser interesante optar por una o otra metodología en función de las circunstancias.

La digitalización manual es mucho más costosa y su resultado es muy variable en cuanto a su precisión espacial, ya que depende en gran medida de la experiencia del operario y de las condiciones de este (cansancio, circunstancias personales, etc.). Por el contrario, e independientemente del operario, el reconocimiento de las entidades es altamente fiable (si se trata de un mapa, este ha sido diseñado para ser interpretado por una persona, por lo que esta reconocerá sus elementos sin dificultad y con total fiabilidad).

Asimismo, un proceso automático, en caso de proceder de forma correcta, tendrá una exactitud absoluta y «clonará» con absoluta fidelidad los elementos del mapa impreso. Esto resulta una ventaja a la hora de obtener una gran precisión, pero impide que en el proceso de digitalización se puedan corregir errores existentes en el documento original. Un operario puede advertir esos errores y corregirlos a medida que digitaliza. Un sistema automático, por el contrario, no puede.

6.5.1. Digitalización manual

La digitalización manual es la forma más básica de crear información digital a partir de un documento cartográfico impreso. Un operario trabaja directamente sobre la fuente cartográfica y su trabajo se traduce en la creación de una nueva capa, gracias a la utilización de un equipo que es capaz de convertir su trabajo en la información necesaria para crear dicha capa.

En el modelo de representación ráster, los elementos básicos son las celdas, que forman una malla regular que puede presentar un numero muy elevado de estas. Una definición manual de las características de cada una de esas celdas resulta inviable, por lo que la digitalización de un documento cartográfico impreso para la obtención de una capa ráster a partir de ella de forma manual no es factible.

Por el contrario, se puede realizar con cierta sencillez la digitalización de una entidad vectorial, trazando la forma de esta o, en caso de ser una entidad de tipo punto, sencillamente indicando su localización. Cuando el número de entidades es elevado, el proceso puede llevar tiempo y ser tedioso, pero en todo caso sigue resultando una forma sencilla y accesible de crear una capa vectorial a partir de otra fuente de datos.

Para llevar a cabo ese trazado de la entidad, se necesita emplear algún equipo que recoja la información introducida por el operador. Existen dos alternativas principales: utilizar un equipo especializado diseñado específicamente para la digitalización, o bien digitalizar utilizando las funciones de edición de un SIG, realizando todo el proceso dentro de este y sin más herramientas que el propio ordenador y un dispositivo señalador como el ratón.

Con equipo especializado (*heads–down*)

La forma tradicional de proceder a la digitalización manual de entidades es utilizando equipos y periféricos expresamente diseñados para llevar a cabo esta tarea. La *tableta digitalizadora* (Figura 6.7) es la herramienta fundamental para este trabajo.

Se trata de una superficie plana a modo de atril, sobre la cual se sitúa el documento cartográfico a digitalizar, y sobre este se van trazando las distintas entidades con un cursor. Este cursor registra los movimientos del operario, convirtiendo las posiciones del cursor en coordenadas reales, que son las que van a constituir la entidad digitalizada. El trabajo del operario consiste en seguir con el cursor las formas de las distintas entidades, como si las estuviera calcando, de modo que indique al sistema las geometrías que se quieren definir.

El proceso de digitalización implica los siguientes pasos [11]:

- **Registro**. La etapa fundamental del proceso, que garantiza que las coordenadas de las entidades digitalizadas sean correctas. El mapa se ha de adherir a la tableta de modo firme, normalmente con cinta adhesiva u otro medio similar, y señalar en él unos *puntos de control* de coordenadas conocidas. Será en base a estos como se calcularan las restantes coordenadas de las entidades que el operario defina mediante el cursor. Habitualmente se utilizan como puntos de control las esquinas y algún punto central del mapa. Es importante que en el proceso de registro el mapa no presente dobleces o deterioros que puedan inducir errores en el cálculo de coordenadas posteriores.
- **Digitalización**. De entidades puntuales, lineales y poligonales.
- **Asignación de atributos**. A cada una de las entidades digitalizadas se le añaden sus correspondientes propiedades. Este paso no se realiza ya con la tableta digitalizadora. En el caso más general, estos atributos se introducen manualmente con el teclado o se toman, por ejemplo, de una base de datos. Un caso particular, no obstante, es el de la digitalización de curvas de nivel. Una vez que estas han sido digitalizadas, no es necesario asignar valores individualmente a cada una de las líneas, ya que entre ellas existe una relación que puede aprovecharse para simplificar el establecimiento de una cota correspondiente a cada una. Estableciendo la elevación de una y la dirección en que la elevación aumenta, pueden sistemáticamente asignarse elevaciones a las curvas que aparecen según se avanza en dicha dirección. Los SIG más populares presentan habitualmente herramientas que facilitan este proceso.

Esta forma de digitalizar se conoce como «cabeza abajo» (*heads–down*), en referencia a la posición del operario a la hora de trabajar sobre la tableta.

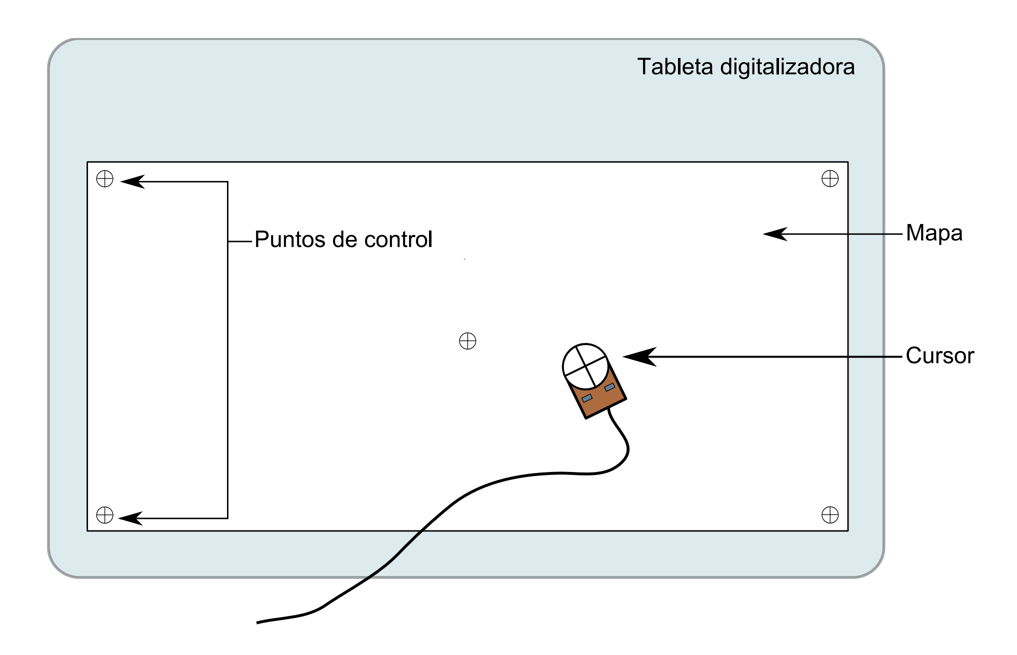


Figura 6.7: Esquema de una tableta digitalizadora y los elementos del proceso de digitalización.

Se distinguen dos formas principales de registro de puntos:

- **Manual**. El usuario debe ir marcando uno por uno todos los puntos que desee incorporar a la entidad digitalizada. Por ejemplo, para el caso de una línea, debe ir deteniendo el ratón regularmente en aquellos puntos que considere de interés, y sobre ellos pulsando los botones del cursor para indicar al sistema que ha de registrar dichos puntos.
- **Semiautomática**. El operario simplemente desliza el cursor definiendo la forma de las entidades, y el propio sistema se encarga de almacenar puntos regularmente según un intervalo de tiempo definido. Esto permite un ahorro de tiempo considerable y una correcta densidad de puntos recogidos para cada entidad.

Las tabletas digitalizadoras son elementos caros, motivo por el cual se tiende a favorecer en la actualidad la digitalización en pantalla, que presenta además otra serie de ventajas adicionales, como seguidamente veremos.

En pantalla (*heads–up*)

La otra forma de digitalizar elementos es utilizando las capacidades de edición de un SIG. Estas capacidades son heredadas de las aplicaciones de diseño asistido por ordenador (CAD), y permiten «dibujar» en la pantalla del ordenador entidades y formas tales como los puntos, líneas y rectas que constituyen los objetos en el modelo de representación vectorial.

En este proceso se parte igualmente de una capa base, generalmente una imagen, y basándose en ella se van definiendo los objetos, «dibujándolos» sobre la pantalla, una vez más como si se calcara aquello que puede visualizarse en dicha imagen. El hecho de que un SIG nos permita tener varias capas simultáneamente y visualizarlas a voluntad facilita el proceso de digitalización. También lo facilita el poder tener varias imáge-



Figura 6.8: Digitalización en pantalla. En rojo, polígono ya digitalizado. Las líneas rojas indican un nuevo polígono, actualmente en edición

nes sobre el fondo (cada una de ellas como una capa individual), de modo que podemos cubrir un área más amplia que la de una simple hoja de mapa o una única imagen.

En este proceso, no partimos en realidad de un documento cartográfico analógico, pues ya ha sido necesario digitalizarlo de alguna forma para incorporarlo en un SIG. El proceso es una digitalización de las entidades como tales, pero la información ya ha de estar en formato digital, aunque no en el modelo de representación vectorial, sino en el modelo ráster. Por ello, puede utilizarse como capa de partida una imagen originalmente en formato digital o bien una imagen originalmente en formato impreso. En este ultimo caso, la imagen ha debido digitalizarse previamente mediante un proceso de *escaneo*, el cual se tratará en la siguiente sección.

En la figura 6.8 puede verse un ejemplo de la digitalización de una imagen en pantalla. Sobre una imagen aérea en color, se digitalizan las distintas parcelas que pueden distinguirse en esta. Del mismo modo, pueden digitalizarse curvas de nivel en un mapa escaneado, u otras entidades tales como ríos, lagos o vías de comunicación sobre una fotografía aérea, entre muchas otras.

La digitalización en pantalla se conoce también como digitalización «cabeza arriba» (*heads–up*), ya que el operador centra su atención en la pantalla, con una postura bien distinta a la que se tiene al trabajar con una tableta digitalizadora.

Frente a dicho trabajo con tableta digitalizadora, la digitalización en pantalla tiene las siguientes ventajas:

- **Menor coste**. No se requiere equipo especializado de alto coste, ya que basta con un ordenador personal.
- **Posibilidad de dividir el trabajo**. Cuando se trabaja con un mapa sobre una tableta digitalizadora, este mapa no puede ser utilizado por otro operario. Sin embargo, el uso de una capa digital dentro de un SIG como base para la digitalización permite que varios operarios trabajen con ella simultáneamente y se repartan el trabajo.
- **Posibilidad de corrección y edición precisa**. Las mismas capacidades que se usan para trazar las distintas entidades pueden emplearse para corregir o modificar estas una vez que han sido digitalizadas (Figura 6.9), resultando esto en un proceso de digitalización más flexible.
- **Posibilidad de ampliación**. Para cartografías de baja calidad, puede ser difícil obtener pre-

cisión si se trabaja directamente sobre el mapa, así como si los elementos a digitalizar son pequeños, requiriéndose del operador un esfuerzo visual adicional. Las capacidades que tiene todo SIG para ampliar una imagen (*zoom*) permiten superar esta dificultad y trabajar a distintas escalas según la precisión del trabajo a realizar o las características de los objetos digitalizados.

- **Mayor precisión**. La capacidad de resolución del ojo humano es mucho menor que la resolución de las imágenes (véase más adelante el apartado 6.5.5). Esto, unido a lo mencionado en el punto anterior, permite aprovechar mejor la información de la fuente original, y que los resultados obtenidos en la digitalización de esta sean más fieles a ella.
- **Mayor comodidad para el operario**. La postura del operario es más adecuada cuando se digitaliza sobre la pantalla, permitiendo unas mejores condiciones. Esto que se traduce en menor cansancio y ello indirectamente comporta resultados más precisos.

Para conocer con más detalle las capacidades básicas de edición de un SIG, así como las restantes capacidades que contribuyen a su vez a facilitar la labor de edición, consúltese el capitulo 23.

6.5.2. Digitalización automática

La digitalización automática limita el trabajo del operario, ya que este no es responsable directo de definir las propiedades de los elementos que se digitalizan. Este tipo de digitalización es la habitual en el caso de generar una capa ráster, aunque también pueden obtenerse capas vectoriales procesando de modo automático cartografía impresa.

Este segundo caso, no obstante, requiere una cartografía en condiciones especiales, no siendo adecuada para todo tipo de mapas. En caso de no presentarse esas condiciones, los resultados de la digitalización no son óptimos, y requieren posteriormente un gran trabajo de corrección y supervisión.

Escaneo

El escaneo es el proceso de digitalización que convierte una imagen impresa (analógica) en una imagen digital [41]. El resultado de este proceso es, por tanto, y desde el punto de vista de un SIG, una capa ráster. Pueden escanearse tanto mapas como fotografías aéreas, operando en ambos casos de un modo similar y con las mismas consideraciones, pues el objeto del proceso es el mismo: la conversión del documento impreso en un documento digital que pueda utilizarse dentro de un SIG o cualquier otro software tal como, por ejemplo, un software de tratamiento de imágenes.

El dispositivo fundamental para realizar este proceso es el *escáner*. Este se compone de una *cabeza* sobre la que se monta un sensor, y un soporte sobre el que se desplaza o bien la cabeza o bien el documento a escanear, de tal modo que durante el proceso de escaneo esta recorre todo el documento, recogiendo la información de toda su extensión.

Este proceso de *barrido* se realiza en una única ocasión, aunque dispositivos más antiguos pueden hacerlo en tres ocasiones a la hora de escanear documentos en color. Aunque lo habitual es la creación de una imagen en color, también pueden obtenerse imágenes en blanco y negro o en escala de grises.

Existen escáneres específicamente diseñados para el trabajo con documentos cartográficos, pero son dispositivos muy especializados y de coste elevado. Los escáneres más genéricos, pensados para el trabajo con todo tipo de imágenes y para todo tipo de usos, pueden no obstante emplearse de igual modo para escanear tanto mapas como



Figura 6.9: Corrección de entidades con las funciones de edición de un SIG. El polígono de la derecha se encuentra en edición, siendo modificado uno de sus vértices.

imágenes aéreas con resultados aceptables, utilizándose con frecuencia.

Existen tres tipos principales de escáneres:

- **De sobremesa** (*flat–bed*). Los habituales para el uso doméstico o el escaneo de imágenes de pequeño formato, aunque también existen de mayor tamaño. El documento a escanear se sitúa sobre una placa de cristal bajo la que se desplaza la cabeza con el sensor. Puede verse uno de estos escáneres en la figura 6.10.
- **De tambor**. El mapa se sitúa sobre un tambor que rota, mientras que la cabeza se mantiene fija. La figura 6.11 muestra uno de estos escáneres.
- **Alimentados**. El sensor se mantiene fijo y el documento se desplaza mediante un mecanismo de arrastre, de forma similar a como avanza el papel en una impresora doméstica. Salvo que dispongan de mecanismos específicos para corregir esta circunstancia, suelen presentar importantes distorsiones geométricas causadas por un desplazamiento impreciso del papel.

Los parámetros básicos que definen las características de un escáner son la resolución espacial y la resolución radiométrica. La primera de estas se mide habitualmente en *puntos por pulgada*[17] y nos indica el número de puntos (celdas) que el sensor es capaz de tomar por cada unidad de longitud sobre el papel. La resolución radiométrica, por su parte, indica la capacidad del sensor para distinguir entre dos colores distintos.

A la hora de trabajar con documentos cartográficos de cara a su posterior utilización en un SIG, tanto la resolución espacial como la radiométrica de los escáneres habituales es en general más que suficiente, incluso en ocasiones en aquellos de uso doméstico. No obstante, es habitual que se presenten distorsiones geométricas que suponen un problema importante a la hora de mantener la precisión cartográfica, y ello exige la utilización de equipos de mayor calidad si se requiere un resultado de alta precisión. Estos equipos no han de ser necesariamente de aquellos pensados para el trabajo con cartografía, sino que pueden ser de uso genérico, siempre, eso sí, que sean de la calidad necesaria.

La velocidad del escáner es otro parámetro importante, pues la preparación de una base de datos cartográfica a partir de cartografía analógica puede llevar un tiempo considerable si el volumen de datos es elevado. El rendimiento del

[17] *Dots per inch*(dpi)



Figura 6.10: Escáner de sobremesa (tomado de Wikipedia)



Figura 6.11: Escáner de tambor (fotografía: Stefan Kuehn)

escáner y la velocidad a la que puede digitalizar una imagen dada está en relación directa con la resolución espacial. Un escáner posee una resolución nominal (en dpi), que es la resolución máxima a la que puede trabajar (el detalle máximo que puede recoger). No obstante, puede ajustarse la resolución de trabajo en función de las necesidades, y una resolución mayor siempre lleva asociado un tiempo de proceso mayor, ya que el volumen de información generado es mayor, así como el detalle que ha de registrarse.

Para cada documento existe una resolución óptima de escaneo en función de las características de este. Esta resolución debe elegirse teniendo en cuenta que el volumen de datos aumenta a medida que empleamos una mayor resolución, buscando un equilibrio adecuado entre ese volumen de datos resultante y la cantidad de información que recogemos. Asimismo, se ha considerar igualmente el tiempo necesario para escanear el documento, tal como se dijo anteriormente.

El parámetro base es la relación entre el tamaño de píxel (la longitud real que representa el ancho de un píxel sobre el terreno) y el tamaño de este píxel en la imagen (lo que mide esa longitud en el mapa). Las resoluciones habituales utilizadas para el escaneo de fotografías aéreas varían entre los 100 dpi ($\approx 250\mu m$ cada punto sobre el mapa) y 2500 dpi (($\approx 10\mu m$ cada punto sobre el mapa) [47]. Por ejemplo para una resolución de 300 dpi, se tiene:

$$\begin{aligned} 300 \text{ dpi} &= \frac{300 \text{ filas}}{2,54 \text{ cm de mapa}} \\ &= 118,11 \text{ filas/cm} \end{aligned} \tag{6.5.1}$$

En un centímetro cuadrado se tienen $118,11^2 \approx 13950$ puntos.

Si trabajamos, por ejemplo, con un mapa a una escala 1:50000, tenemos que la distancia real que representa el alto de cada fila es

$$\frac{50000 \text{ cm}}{118,11 \text{filas}} = 4,24 \text{ metros/fila} \tag{6.5.2}$$

Es decir, cada píxel del mapa representa sobre el terreno un cuadrado de lado 4,24 metros.

Con cálculos similares podemos calcular para cada posible resolución el espacio real que representa, y elegir esta en función del detalle que necesitemos. Como regla general, debe tratar de

trabajarse con una resolución que garantice que los objetos que resultan de interés de la imagen (por ejemplo, aquellos que van a digitalizarse después manualmente mediante una digitalización en pantalla con esa imagen) sean distinguibles con claridad.

En el caso de imágenes aéreas, la resolución de estas medida en pares de líneas por milímetro puede ser superior y permitir escanear a mayor resolución, aunque ello no es estrictamente necesario, y debe una vez más buscarse el equilibrio entre las ventajas y los inconvenientes de trabajar con una resolución más elevada.

En [47] puede encontrarse información más detallada sobre la elección de una resolución óptima en el escaneo de imágenes aéreas.

Para el caso de mapas, no deben olvidarse los fundamentos cartográficos en base a los cuales se ha creado dicho mapa, que fueron detallados en el capítulo 3. Trabajar con una resolución más elevada no hace necesariamente que estemos incorporando más información, ya que esta puede no existir en el mapa original. Tendríamos un volumen de datos más elevado que el necesario para recoger toda la información del mapa.

Una diferencia fundamental entre escanear una hoja de un mapa y una imagen aérea es la diferencia de tamaño. Los mapas suelen tener tamaños mucho mayores que los de un escáner común, lo cual obliga a utilizar equipos de gran formato o, en la mayoría de los casos, contratar servicios de escaneo especializados, ya que estos equipos tiene un coste muy elevado.

Una solución distinta en el caso de mapas de gran tamaño es el escaneo de la hoja por partes y la posterior unión de las distintas partes. En este caso, es necesario asegurarse de que las partes son coherentes entre sí en lo que respecta a las condiciones bajo las que se realiza el escaneo, así como garantizar que las distintas partes se solapan para que no existan zonas sin datos en la imagen resultante. Además de esto, el solape facilita la localización de puntos comunes presentes entre partes contiguas, lo que ayuda en la composición de todas las partes para dar lugar al resultado global.

Otra diferencia entre trabajar con mapas e imágenes es la relativa al tipo de soporte. En el caso de mapas, el documento original se encuentra siempre impreso en papel. En el caso de fotografías aéreas puede presentarse tanto en papel como en diapositiva. Los escáneres están preparados para capturar la imagen tanto por *reflexión* (cuando se trabaja con un documento en papel) como por *transmisión* (cuando se trabaja con una diapositiva o cualquier otro soporte transparente), por lo que ambos tipos de fuentes pueden utilizarse indistintamente para generar una imagen digital, siendo esta diferencia menos relevante a efectos prácticos.

Por último, un aspecto clave en el escaneo de cartografía es la asignación de coordenadas a la capa resultante. Cuando utilizamos una tableta digitalizadora, debemos definir los *puntos de control*, que son los que establecen la referencia geográfica en base a la cual se calculan las coordenadas de los elementos que digitalizamos con el cursor. En el caso de escanear un mapa o una fotografía aérea, esa información está presente en el mapa en forma de marcas fiduciales o una retícula con coordenadas impresas, pero no se digitaliza como tal.

Si simplemente escaneamos el documento, se digitaliza la marca fiducial o la etiqueta que indica las coordenadas, pero tan solo como una imagen, y no como un dato aprovechable por el SIG para otras tareas. En esta imagen, un operador puede ver las coordenadas de un punto, pero si realizamos un proceso de digitalización vectorial en pantalla utilizando esa imagen, el SIG no tiene forma de calcular las coordenadas de

los puntos que introducimos, pues carece de una referencia.

Para que una imagen procedente del escaneo de un documento impreso tenga plena validez y utilidad dentro de un SIG, es necesario añadirle información sobre la localización en el espacio del área representada en dicho documento. Este proceso se denomina *georreferenciación*.

La georreferenciación es un proceso tratado dentro de este libro en el apartado 17.2.1, puesto que no es puramente un proceso que forme parte de la adquisición de datos, sino un tratamiento a aplicar una vez que el proceso de digitalización ha sido realizado.

Vectorización automática

La vectorización automática es un proceso completamente distinto al de escaneo, y no es tan habitual en el ámbito de los SIG, principalmente debido a la mayor dificultad que entraña. Como resultado de este proceso, se obtiene una capa vectorial, pero, a diferencia de la vectorización manual, el operario no tiene que señalar los puntos de estas o trazar los contornos de las entidades.

Existen distintos procesos de vectorización automática, entre los que distinguiremos los siguientes:

- Vectorización en base a una imagen digital, por reconocimiento de entidades en un software apropiado.
- Vectorización mediante dispositivos específicos que trabajan sobre un documento analógico.

En el primer caso, partimos de una imagen digital, que puede proceder o no de un proceso de escaneo. Sobre esta imagen se aplican algoritmos que identifican de modo automático las distintas entidades y crean los correspondientes objetos vectoriales.

El mayor inconveniente de esta técnica es que requiere que la imagen tenga unas condiciones especiales, pues de otro modo es difícil que esos algoritmos de identificación den resultados correctos. En ocasiones pueden crear entidades donde estas no existen o bien ignorar algunas por no ser capaces de detectarlas, así como crear entidades de forma y tamaño incorrectos. El trabajo de digitalización por parte del operario desaparece, pero es necesario un trabajo posterior de comprobación y corrección, que en función de las características de la imagen de partida puede ser importante.

Esta forma de vectorización automática es, al igual que la georreferenciación, un proceso a llevar a cabo sobre la imagen. Por esta razón, no se trata en este capítulo sino en el capítulo 17 dedicado al tratamiento de imágenes. Igualmente, el capítulo 18, dedicado a la conversión entre capas ráster y vectoriales, incluye información acerca de procesos de vectorización automática, con particular atención a la conversión de un mapa escaneado en una capa vectorial de curvas de nivel.

La otra forma de digitalización es totalmente diferente y no se realiza en el ordenador, sino en un periférico externo a este, tal como una tableta digitalizadora o un escáner. El dispositivo en cuestión es más similar a un escáner que a una tableta digitalizadora, pero su comportamiento imita al de un operario trabajando sobre esta última.

Para ello, dispone de sensores luminosos y de láser que buscan las líneas en la imagen y las recorren, almacenando las coordenadas por las que han pasado en el recorrido. De este modo, se genera un resultado vectorial en lugar de uno ráster. El barrido de la imagen no es sistemático como el de un escáner, sino que «sigue» las líneas que están presentes en la imagen, y que son las que van a digitalizarse.

Al igual que con la digitalización automática, las buenas condiciones de la imagen de partida son básicas para obtener resultados de calidad. En un mapa es habitual que las líneas se vean interrumpidas por etiquetas (por ejemplo, para indicar la altura de una curva de nivel), o bien están dibujadas en trazo punteado, o bien puede aparecer alguna mancha sobre ellas. Este tipo de elementos dificultan o incluso imposibilitan el correcto funcionamiento del dispositivo, ya que este no puede seguir las líneas adecuadamente, lo cual da lugar a resultados de poca calidad.

6.5.3. Digitalización a partir de coordenadas. Geocodificación

Junto a las formas de digitalización que acabamos de ver, existe una forma aún más básica: la digitalización directa de valores y coordenadas, sin necesidad alguna de dispositivos especializados o elementos gráficos. En este tipo de digitalización no existe un mapa o documento cartográfico, sino simplemente una serie de datos espaciales expresados de forma alfanumérica que son susceptibles de convertirse en una capa y emplearse así dentro de un SIG.

Este proceso se conoce como *geocodificación* [48] e implica la asignación de coordenadas a puntos de interés, los cuales pueden ser de naturaleza muy variada. Asimismo, la procedencia de estos datos también puede ser diversa, y en general muchas formas de trabajo en campo dan lugar a datos que, aún no estando originalmente dispuestos sobre mapas, sí que pueden emplearse como base para la creación de capas. Algunos ejemplos son los siguientes:

- Muestreos de campo tales como la medición de parcelas en un inventario forestal. Cada parcela tiene una coordenada correspondiente a su centro, y los árboles medidos se referencian con un rumbo y una dirección en base a ese centro.
- Calicatas para análisis de suelo
- Levantamientos topográficos con instrumentación tanto analógica como digital. Existe un conjunto de instrucciones y procedimientos denominado COGO (*COordinate GeOmetry*), que facilita el trabajo con datos en forma de distancias y ángulos, de forma que las mediciones efectuadas a lo largo de un recorrido empleando un equipo tal como una estación total, un teodolito o un nivel con una mira, todos ellos pueden posteriormente convertirse con sencillez a coordenadas mediante la incorporación al SIG de ese conjunto de valores.
- Coordenadas en las que han sucedido algún tipo de sucesos. Por ejemplo, la geocodificación de localizaciones en las que han tenido lugar sucesos criminales permite posteriormente el análisis de su distribución y el establecimiento de políticas de seguridad más acordes con el escenario real.
- Coordenadas de cierto tipo particular de elementos, tales como elementos arquitectónicos, árboles singulares, paradas de autobús. Estas permiten la localización rápida de estos y una fácil catalogación, además de, en conexión con otras capas, cálculos como, por ejemplo, la forma más rápida de desplazamiento hasta uno de ellos.
- Coordenadas correspondientes a otras formas de codificación espacial. Sistemas de localización espacial tales como códigos postales o, por ejemplo, los sistemas de indexación espacial CGDG o *c-squares*, pueden todos ellos vincularse a coordenadas geográficas, de tal modo que a cada uno de los códigos de estos sistemas se le asigne una de tales coordenadas.
- En la actualidad, Internet está viendo aparecer tendencias relacionadas con la asignación de una localización geográfica a muchos de sus elementos. Así, puede añadirse a una página Web información sobre el emplazamiento donde ha sido creada, o añadirla a una fotografía digital. Los

datos con los que trabajamos en la Web (textos, imágenes, etc.) llevan asociados a su vez otros datos (metadatos) con información sobre su localización. El proceso de añadir estos metadatos se conoce como *geotagging*.

Todos estos datos presentan en común que, recogidos de un modo u otro, conforman un conjunto de coordenadas puntuales que habitualmente sirven para el trabajo fuera de un SIG y no llegan a incorporarse a este, o que al menos no están dispuestos en la forma habitual de capa con la que trabajamos en un SIG.

En el caso de encontrarse en formato analógico, estos datos pueden digitalizarse mediante la simple introducción manual de coordenadas a través del teclado o bien mediante algún sistema más específico como el escaneo del documento y el empleo de algún software de reconocimiento de caracteres (OCR)[18].

En el caso de encontrarse ya en formato digital, estos datos pueden presentarse como tablas en una hoja de cálculo, datos asociados a otro dato de cualquier tipo (como en el caso del *geotagging*) o incluso simples archivo de texto. Muchos SIG incorporan métodos para leer estos archivos y después utilizar las coordenadas que contienen con el fin de crear una nueva capa, en general de puntos.

Un caso particular de la creación de puntos con coordenadas es la asignación de direcciones dentro de núcleos urbanos, tales como direcciones postales o códigos postales. Estas direcciones son de especial importancia en el desarrollo de actividades dentro del entorno urbano, ya que es más habitual referirse al emplazamiento de un determinado elemento (por ejemplo, un comercio) en términos de su dirección postal que en coordenadas espaciales tales como las que se manejan en un SIG. La geocodificación de estos elementos implica establecer una coordenada geográfica correspondiente a cada dirección postal.

[18]Optical Character Recognition

6.5.4. Fotogrametría

Un caso particular de digitalización lo encontramos en la *fotogrametría*. En la definición clásica de [49], esta se define como la técnica para estudiar y definir con precisión la forma, dimensiones y posición en el espacio de un objeto cualquiera, utilizando medidas realizadas sobre una o varias fotografías. Esta definición no limita el alcance de la fotogrametría al ámbito de lo geográfico, y se utilizan sus principios en campos tales como la arqueología o la documentación de obras y monumentos, empleando para ello fotografías no aéreas, sino terrestres. Es la denominada *fotogrametría terrestre*. No obstante, la rama de interés para este libro es la de la *fotogrametría aérea*, cuya base de trabajo tradicional son las fotografías aéreas.

Esta clase de fotogrametría viene, pues, ligada íntimamente a los inicios de la teledetección, cuando los sensores modernos que hemos estudiado antes en este mismo capítulo no se habían desarrollado, y los existentes (básicamente cámaras fotográficas especialmente adaptadas a la toma de fotografías de tipo cartográfico) se montaban a bordo de aviones.

Históricamente, el término *teledetección* aparece con posterioridad, una vez que las técnicas de toma de imágenes avanzan y dan un gran salto cualitativo con la aparición de las imágenes satelitales y los sensores electro–ópticos que ya conocemos. Algunos autores engloban la fotogrametría dentro de la teledetección, mientras que otros se refieren con el termino teledetección a las tecnologías más actuales y las consideran disciplinas distintas aunque muy relacionadas. Junto con la fotogrametria aérea aparece la fotograme-

tría espacial, encargada de operar sobre imágenes de satélite bajo unos principios similares.

Dentro de este libro entenderemos por teledetección todo el conjunto de técnicas y operaciones de obtención de imágenes (que ya conocemos), así como las de tratamiento y posterior extracción de resultados a partir de estas (que iremos viendo en otros capítulos), obteniéndose estos resultados sin necesidad de establecer contactos con los objetos a estudiar, como corresponde a la definición dada en el apartado correspondiente. Dentro de ese conjunto de operaciones que nos llevan desde las imágenes a los resultados, entendemos como parte de la fotogrametría aquellas que tienen relación con la acepción original del término, es decir, aquellas que derivan de la medición de elementos.

La denominación, no obstante, no es tan relevante, y sí lo es sin embargo comprender la importancia de ambas, particularmente dentro de este capítulo como técnicas de producción cartográfica.

En lo que respecta a la fotogrametría, el proceso de *restitución* es el que interesa principalmente para el contenido de este capítulo, pues ofrece como resultado nuevas capas de datos tanto bidimensionales como, especialmente, tridimensionales. Así, pueden obtenerse tanto las capas vectoriales digitalizadas que veíamos por ejemplo en el apartado 6.5.1, como directamente Modelos Digitales de Elevaciones a partir de imágenes.

En realidad, los procesos de digitalización que ya hemos visto son también parte de la fotogrametría digital, y es habitual encontrarlos en los textos al uso sobre esta. También lo son los procesos de rectificación que se han citado en su momento, y que analizaremos en detalle más adelante en el capítulo 17. Como puedes ver, todas las técnicas están sumamente relacionadas, y las divisiones que hacemos pueden ser unas u otras en función del enfoque que se dé para su estudio

Todas estas operaciones se llevan a cabo con una *estación fotogramétrica*, que comprende las herramientas necesarias para llevar estas a cabo (algunas, como los escáneres, ya las conocemos). En función del tipo de herramientas y técnicas distinguimos los siguientes tipos de fotogrametría, que representan a su vez la evolución de la disciplina.

- Fotogrametría **analógica**. Basada en mediciones y procedimientos sobre imágenes analógicas
- Fotogrametría **analítica**. Basada en formulaciones matemáticas y técnicas computacionales, permite obtener grandes precisiones.
- Fotogrametría **digital**. Basada en el trabajo con imágenes digitales dentro de un entorno informático.

El interés principal desde el punto de vista de los SIG es en la fotogrametría digital, ya que existe una gran relación entre estos y las aplicaciones empleadas en dicho tipo de fotogrametría. Es en esta en la que pueden englobarse los procesos de digitalización que ya hemos visto, y no en las restantes formas más antiguas de fotogrametría. En la fotogrametría digital, la estación fotogramétrica se articula sobre un ordenador en el cual se llevan a cabo los distintos procesos, no existiendo operaciones externas al mismo. Así, las imágenes se manejan dentro del ordenador y se visualizan a través de él, y la generación de nueva cartografía también se produce de forma digital.

Esto no es muy diferente de lo que veíamos en el caso de la digitalización en pantalla algunas paginas atrás, pero el trabajo fotogramétrico engloba otros procesos además de los que ya hemos visto. Uno de ellos es la generación directa de cartografía de elevaciones, para la cual se requiere que el equipo empleado disponga de algunos elementos adicionales. Es decir, la estación foto-

gramétrica digital es más compleja que un simple ordenador, un dispositivo de marcado (un ratón) y un SIG, que eran los requisitos básicos para digitalizar en pantalla una imagen.

Una estación fotogramétrica digital ha de tener, por ejemplo, capacidad para generar visualizaciones con sensación de profundidad a partir de pares de imágenes, que son las que permiten la posterior digitalización de los elementos con sus elevaciones correspondientes. Los principios en los que se basan este tipo de visualizaciones son los mismos empleados en la fotogrametría no digital, fundamentados en la visión estereoscópica.

La visión tridimensional en el ser humano se basa en el hecho de que la imagen que ve cada ojo es ligeramente distinta a la del otro, lo cual permite al cerebro extraer información volumétrica y generar una verdadera visión tridimensional. En el caso de la fotogrametría, si en lugar de utilizar una única imagen aérea o de satélite empleamos dos, cada una de ellas tomada desde un punto distinto, resulta posible recrear el efecto que ambas imágenes tendrían para la reconstrucción tridimensional de la escena, y «engañar» al cerebro del observador para que este pueda observar la escena con volumen y profundidad.

Cuando se emplean imágenes de satélite, los pares se pueden obtener con aquellas plataformas y sensores que permiten variar el ángulo de visión, de modo que en la misma pasada del satélite se toman imágenes de una zona desde distintos puntos. El sensor toma una imagen cenital y posteriormente, una vez ha superado la zona en su recorrido, toma una segunda imagen mirando «hacia atrás», la cual, combinada con la primera, permite el levantamiento del terreno y la realización de los procesos fotogramétricos (Figura 6.12).

El sensor HRS que montan los satélites SPOT, o el sensor ASTER, ambos son capaces de tomar este tipo de imágenes.

Las formas de conseguir que el observador perciba la profundidad de la escena a partir de las imágenes son variadas, y van desde el uso de sencillos instrumentos ópticos o la generación de anaglifos (imágenes que combinan la información del par estereoscópico y que se han de observar con gafas con filtros distintos para cada ojo), hasta otras técnicas más complejas y elaboradas. En la fotogrametría no digital, el empleo de restituidores analíticos ha sido la metodología habitual. En la fotogrametría digital, este puede sustituirse por un equipo con dos monitores, cada uno de los cuales muestra una de las imágenes del par, y se emplean gafas especiales que son las encargadas de generar en el observador la sensación de profundidad .

Además de lo anterior, la estación fotogramétrica digital dispone de periféricos específicos tales como ratones 3D, o manivelas como las que presentan los restituidores analíticos, facilitando así la adaptación de los operarios a este tipo de estación.

Por último el software que implementan, y que es el encargado de representar las imágenes y acoger el proceso de digitalización, suele ser específico, y es frecuente que se distribuya como parte de toda una estación fotogramétrica compuesta por los elementos reseñados anteriormente. Algunos SIG incorporan progresivamente capacidades adaptadas de este tipo de programas, pero por el momento la labor fotogramétrica queda reservada para este tipo de aplicaciones específicas, siendo el SIG tan solo un beneficiario directo de sus productos.

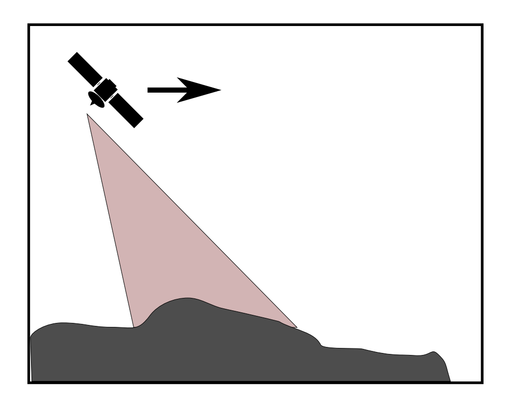
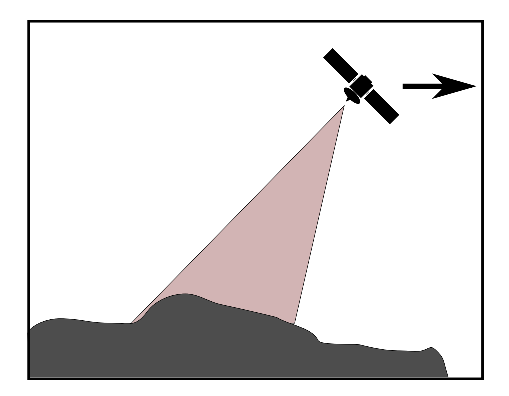



Figura 6.12: Toma de pares de imágenes estereosópicas desde un satélite, mediante variación del ángulo de visión.

6.5.5. Calidad de la digitalización

Uno de los aspectos más importantes del proceso de digitalización es la calidad del resultado obtenido, que debe tratar de ser lo más cercana posible a la calidad original de la información que se digitaliza, es decir, del mapa o imagen original. Independientemente de la precisión del equipo utilizado o la habilidad y experiencia del operario, la digitalización no es por completo perfecta, conteniendo siempre ciertas deficiencias y errores.

Además de los errores que puedan incorporarse en las distintas fases del proceso de digitalización (sea este del tipo que sea), hay que considerar que las fuentes originales a digitalizar también pueden incluir los suyos propios. Así, el proceso de escaneado puede incorporar distorsiones geométricas, pero es posible que el mapa o fotografía aérea de partida también presente alguna distorsión como consecuencia de su deterioro, más patente cuanto más antigua sea esta.

La información contenida en el documento cartográfico puede también contener elementos problemáticos de cara a obtener un producto de calidad, que pueden ir desde líneas borradas total o parcialmente a manchas en el propio mapa derivadas de su uso habitual [11].

Dentro de los errores que aparecen como consecuencia de la digitalización en sí, un tipo importante de ellos son las discrepancias y coincidencias imperfectas entre las distintas entidades, tal como las que se muestran en la figura 6.13

Estas imprecisiones son causantes de numerosos problemas, tales como la aparición de polígonos espúreos en las operaciones de solape entre capas vectoriales, que veremos en el capítulo 19.

Debido a esto, las capacidades de edición de los SIG incorporan funcionalidades que permiten evitar estos errores en el momento de la digitalización, ayudando al operario en su tarea y permitiéndole alcanzar una exactitud y precisión imposible de lograr sin estas funcionalidades. Entre ellas, es especialmente importante el establecimiento de tolerancias y ajuste automático en función de ellas (esto se conoce con el término ingles *snapping*), que ayudan a garantizar la coincidencia entre los distintos vértices.

De este modo, polígonos adyacentes o líneas que se cortan en un punto dado lo hacen con total exactitud. Dichos polígonos comparten exactamente el mismo lado con las mismas coordenadas exactas, o se cruzan en el mismo e idéntico punto, y no únicamente pasan por un punto cercano (pe-

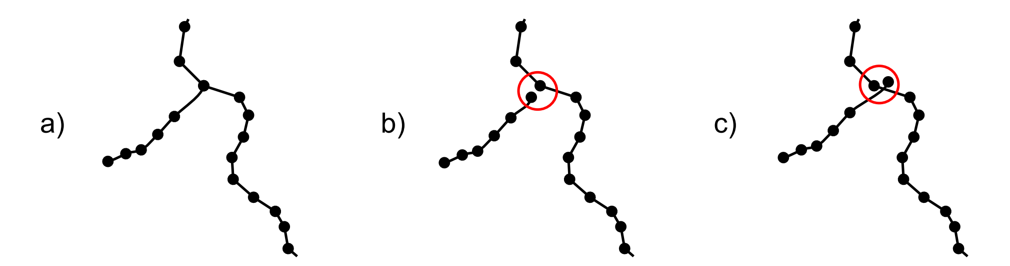


Figura 6.13: Errores derivados del proceso de digitalización. a) Versión correcta, con nodos coincidentes. b) y c) Versiones con errores que causan una falsa desconexión entre las líneas.

ro distinto) definido con la precisión con la que el operador haya podido ajustar ambas entidades visualmente. La coincidencia no es solo visual, sino numérica. La figura 6.14 muestra un ejemplo de la utilización de *snapping* en un proceso de digitalización.

Mediante estas funcionalidades, el operador simplemente selecciona un punto, y el sistema digitalizador lo desplaza para que coincida con el punto existente más cercano, siempre que se encuentre a menos distancia que la tolerancia establecida de antemano.

El hecho de que exista una completa coincidencia es especialmente importante cuando la capa vectorial que se digitaliza contiene información topológica. La topología exige que la coincidencia sea correcta y defina perfectamente la relación entre las entidades. Para los ejemplos b) y c) de la figura 6.13, las líneas no están conectadas ya que no existe coincidencia en el nodo. Si los puntos están suficientemente cercanos, puede «parecer» que son coincidentes, pero el SIG no los detectará como tales y no se podrá llevar a cabo ningún análisis topológico con esas líneas (por ejemplo, suponiendo que representan vías de comunicación y se quiere hacer un análisis de redes con ellas).

La digitalización de entidades en caso de querer recoger su topología debe obedecer una serie de reglas, a saber:

- Las líneas deben cruzarse en nodos, en caso de que exista relación (conexión) entre ellas.
- Las líneas que coinciden en un nodo común deben coincidir exactamente. Las funciones de *snapping* se han de utilizar por ello durante la digitalización.
- Los lados comunes de los polígonos deben digitalizarse una única vez.
- Las áreas deben ser cerradas (el primer punto ha de coincidir exactamente con el último). Las funciones de *snapping* o el cierre automático de líneas (asignar sistemáticamente al último punto del contorno del polígono las coordenadas del primero) deben emplearse para ello.

Todos los aspectos relativos a la calidad de datos, entre los cuales se incluyen los relacionados con los errores del proceso de digitalización, se tratan con mayor profundidad en el capítulo 7.

6.6. GPS

Uno de los hitos en la aparición de nuevas fuentes de datos geográficos es la aparición de los *Sistemas Globales de Navegación por Satélite*

[19] *Global Navigation Satellite System*

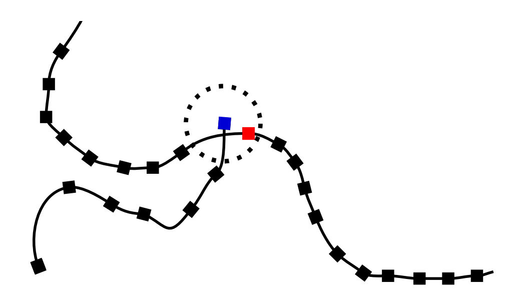


Figura 6.14: Ajuste automático mediante tolerancia(*snapping*). El nodo azul representa el nodo en edición. La tolerancia de enlace queda marcada por el circulo punteado. Puesto que el nodo rojo de la línea preexistente se encuentra dentro de esa tolerancia, al añadir el nuevo nodo (azul), este automáticamente se situará en las coordenadas del nodo rojo, garantizándose así la coincidencia.

(GNSS)[19], que permiten la obtención de coordenadas geográficas de un modo inmediato, con las consecuencias que esto tiene para su uso en actividades como la elaboración de cartografía.

En esencia, un GNSS es un sistema que permite conocer en todo momento y en cualquier punto del globo la localización exacta de dicho punto con un margen de error del orden de unos pocos metros o menos. Para ello, se basan en el envío de señales entre un dispositivo situado en el punto concreto y una red de satélites, pudiendo establecerse la posición exacta mediante las características de dicha transmisión.

El ejemplo más extendido de un GNSS es el Sistema de Posicionamiento Global (Global Positioning System, o GPS)[20], originalmente puesto en funcionamiento por el Departamento de Defensa de los Estados Unidos. Existen otros GNSS operativos tales como el GLONASS ruso, el COMPASS chino o el *Galileo* europeo)

6.6.1. Fundamentos del sistema GPS

El sistema GPS se divide en tres subsistemas o *segmentos*:

- **Segmento espacial**. Lo componen los satélites de la constelación GPS (un total de 27, siendo 24 de ellos operativos y 3 de reserva), con los cuales se comunican las unidades receptoras, y en función de los cuales puede triangularse la posición actual de estas.
- **Segmento de control**. Lo forman un conjunto de estaciones terrestres que controlan el funcionamiento de los satélites, pudiendo enviar señales a estos para modificar su comportamiento.
- **Segmento de usuarios**. Lo conforman los receptores GPS y todos los dispositivos que hacen uso de la señal de los satélites para el cálculo de posiciones.

Los satélites del segmento espacial emiten una señal compleja cuyo contenido puede dividirse esencialmente en dos bloques de información:

- **Señales empleadas para el cálculo de distancias**. Estas incluyen dos códigos: P(Precise) y C/A (Coarse/Aquisition). El segundo de ellos es el empleado habitualmente, ya que el primero se encuentra encriptado y está pensado para uso militar, mientras que el C/A esta disponible para todos los usuarios.
- **Mensajes de navegación**. Estos informan de la posición orbital del satélite (conocida como

[20]El nombre completo del sistema es NAVSTAR–GPS (NAVigation SysTem And Ranging - Global Position System)

efeméride), y pueden asimismo contener información adicional referente al segmento espacial.

Las señales para el cálculo de distancias (en la terminología GPS estas distancias se conocen como *pseudodistancias*) se envían mediante una onda portadora conocida como L1, correspondiente a una frecuencia de 1575,42 MHz . El código P se envía además en una segunda portadora denominada L2, con una frecuencia de 1227,60 MHz.

El funcionamiento del sistema se basa en la triangulación de la posición mediante las señales procedentes de un cierto número de los satélites. Esta posición se calcula no únicamente en sus coordenadas *x* e *y*, sino también en *z*, es decir en elevación. El sistema GPS emplea como sistema geodésico de referencia el WGS84 [50]. La precisión en el cálculo de la elevación es menor que la correspondiente a las restantes coordenadas, aunque también es de utilidad y puede emplearse en aplicaciones que van desde levantamientos y replanteos a usos en tiempo real como el cálculo de elevación en vuelos [51].

La posición de los satélites es conocida en todo momento, y los propios satélites informan de ella a los receptores a través de los mensajes de navegación. En base a esas posiciones orbitales, el proceso de triangulación que se lleva a cabo en el sistema GPS no se basa en el trabajo con ángulos, sino con distancias (aunque es habitual hablar de triangulación, en realidad se trata de una *trilateración*).

El cálculo de la distancia puede realizarse utilizando la información de las señales (los códigos C/A o P), o bien empleando las propias portadoras. El primer método es más sencillo y rápido, ya que no es necesario que el receptor «escuche» la señal durante un periodo prolongado de tiempo, lo cual sí es necesario en el segundo, como a continuación veremos.

En el caso de emplear la portadora, se mide el desfase entre esta y una señal generada por el receptor, lo cual permite calcular una parte de la distancia (la que es menor que la longitud de onda de la señal). La distancia total es igual a esta parte calculada más un numero entero de veces la longitud de onda. El valor de este numero entero es, no obstante, desconocido. Su cálculo se conoce como *resolución de la ambigüedad* (AR), y requiere escuchar la señal del satélite durante un cierto tiempo para recopilar datos suficientes que permitan el cálculo del valor antedicho.

Así, la resolución de la ambigüedad es la que hace necesario un tiempo de *inicialización* de la unidad, con objeto de conocer esa constante en el desfase. Si la unidad pierde contacto con el satélite, es necesario de nuevo proceder a la resolución de las ambigüedades, quedando el receptor inoperativo durante ese periodo de tiempo. Más detalles sobre la resolución de la ambigüedad en el sistema GPS puede encontrarse en [52].

Puesto que la velocidad a la que la señal se desplaza es muy elevada, se requieren relojes muy precisos para poder medir con precisión los tiempos tan cortos que tarda dicha señal en recorrer la distancia entre satélite y receptor. A bordo de los satélites se montan relojes atómicos de muy alta precisión, pero las unidades receptoras no disponen de relojes tan precisos. Es por este motivo que, como veremos, han de introducirse correcciones y cálculos adicionales con el fin de obtener mayores precisiones en la medida del tiempo.

Si el receptor es capaz de establecer comunicación con tres satélites, dispone ya de información suficiente para conocer su posición (x, y) como intersección de las esferas centradas en cada uno de dichos satélites y con radio la distancia existente entre este y el receptor. Con cuatro satélites se puede ya obtener la posición (x, y, z).

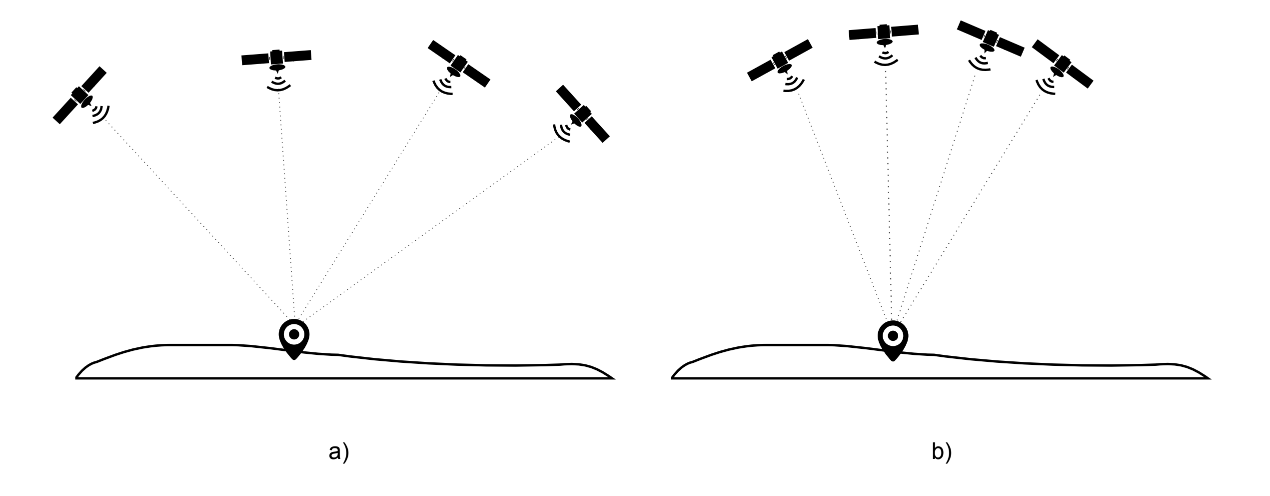


Figura 6.15: Dilución de la precisión. La geometría de los satélites en el ejemplo a) da una mayor precisión en el cálculo de la posición del receptor que la del ejemplo b).

Un número mayor de satélites (cuatro al menos) es necesario, no obstante, para eliminar las imprecisiones debidas a los distintos elementos implicados, y se emplean habitualmente modelos más complejos que utilizan los datos de múltiples satélites y efectúan correcciones en función de ellos. Las deficiencias de los relojes que emplean los receptores pueden corregirse mediante la utilización de nuevos satélites, que permiten calcular con exactitud el tiempo, variable de gran importancia en el proceso y sin la cual no se pueden obtener precisiones elevadas.

Los receptores actuales están preparados para trabajar con un número máximo de satélites habitualmente igual a 12, por lo que en todas circunstancias tratan de localizar el mayor número posible de satélites con objeto de lograr una mayor precisión.

El diseño de la red de satélites está pensado para garantizar que en cualquier punto de la superficie terrestre y en cualquier momento, un receptor puede localizar el número necesario de satélites para obtener con exactitud su precisión. La localización en la que se disponen los satélites con los que se establece comunicación no es irrelevante, ya que condiciona la precisión del posicionamiento, afectando a lo que se conoce como *dilución de la precisión* (DOP[21]). Si los ángulos de los satélites son grandes, la precisión que se obtiene es mayor que si estos son menores (Figura 6.15).

Junto a esto, existen otras muchas fuentes de error en el sistema GPS, cada una de las cuales afecta a la precisión del mismo. Entre ellas, cabe destacar las siguientes:

- Errores en la posición de los satélites.
- Errores por el rebote de la señal en otros elementos tales como edificios, con anterioridad a alcanzar el receptor.
- Errores derivados del paso de la señal por la atmósfera. Al atravesar la ionosfera y la troposfera se genera un retraso por la alteración que dicho paso produce sobre la señal.

[21]Dilution of Precision

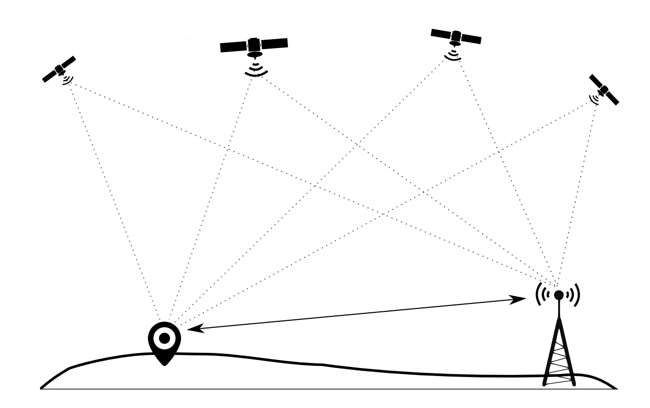


Figura 6.16: Esquema de funcionamiento del GPS diferencial

- Errores en la precisión de los relojes, ya mencionados.
- *Disponibilidad selectiva.* Debido a su concepción como una herramienta militar, el departamento de Defensa de los Estados Unidos, propietario del sistema, introducía errores aleatorios en las señales, de tal forma que esta quedaba degradada y los usuarios civiles no podían obtener una precisión muy elevada. La disponibilidad selectiva fue eliminada en el año 2000.

En conjunto, todos estos errores suman desviaciones apreciables, que sin embargo pueden corregirse con la aplicación de técnicas adicionales, por ejemplo incorporando información adicional procedente de otros receptores. Una de estas técnicas es el denominado *GPS diferencial*, pensado en origen para eliminar el error de la disponibilidad selectiva, aunque también eficaz para corregir una buena parte los restantes errores citados anteriormente.

Para la aplicación del GPS diferencial se requiere no solo un receptor único (aquel del cual se quiere calcular su posición), sino también otro receptor fijo de referencia cuyas coordenadas se conocen con gran precisión. Este receptor fijo es, a su vez, un receptor de alta precisión y, además de calcular su propia posición, emite información que las unidades receptoras pueden aprovechar para corregir sus mediciones. El receptor móvil, lógicamente, tiene que soportar este tipo de correcciones, para poder hacer uso de la señal de la estación de referencia.

Los datos que permiten llevar a cabo la corrección puede obtenerse en el receptor mediante radio, descargarse por Internet mediante una conexión inalámbrica, o bien utilizar una constelación de satélites adicional dedicada a elaborar y servir este tipo de datos.

La corrección puede realizarse fuera del propio receptor, a posteriori, utilizando software adecuado y los mismos datos de corrección que si se realiza la corrección en tiempo real.

El fundamento de este sistema es que los errores que afectan al receptor móvil también afectan al de referencia. La magnitud del error que afecta al receptor de referencia puede conocerse, ya que se conoce la coordenada exacta de este, y en base a eso puede eliminarse el error que afecta al receptor móvil, asumiendo que ambos errores son de similar índole.

En la actualidad, aplicando estas técnicas de corrección diferencial, un GPS puede obtener precisiones del orden de 2 metros en latitud y longitud, y 3 en altitud. Sin corrección diferencial, esta precisión es de unos 10–20 metros.

La figura 6.16 muestra un esquema del funcionamiento del GPS diferencial.

6.6.2. Tipos de receptores

La precisión del sistema global GPS depende del tipo de receptor GPS (o, en el lenguaje común, GPS a secas) que se emplee, obteniéndose mayores precisiones con receptores más avanzados, siempre dentro de las posibilidades del propio sistema.

En función de sus características y de la forma en que operan, podemos distinguir los siguientes tipos de receptores GPS:

- **Receptores secuenciales**. Establece conexiones secuenciales con los distintos satélites disponibles, estando conectado a uno o dos a lo sumo simultáneamente. Estos receptores son más económicos, ya que esta forma de operar requiere equipos menos complejos, aunque la precisión que se obtiene también es menor.
- **Receptores continuos**. Disponen de más canales de radio que los anteriores y ello permite que la conexión a los satélites sea continua, sin tener que alternar entre uno y otro. La precisión que se obtiene es mayor, pero se trata de equipos más caros.
- **Receptores con canales multiplexados**. El esquema de funcionamiento es similar al secuencial, alternando entre los distintos satélites y utilizando un único canal. No obstante, utilizan software más complejo y procesadores más potentes, de forma que esta alternancia se puede producir con una frecuencia mucho más elevada.

A día de hoy, es habitual que incluso los GPS de menor coste tengan múltiples canales, permitiendo la conexión continua con un número elevado de satélites.

Como hemos visto, las señales emitidas por los satélites contienen dos códigos (C/A y P) que se transmiten modulados sobre dos ondas portadoras distintas (L1 y L2). No todos los receptores GPS son capaces de utilizar estos elementos de las señales, y en función de ello podemos también clasificarlos.

Los más sencillos únicamente basan sus cálculos en el código C/A, mientras que los más avanzados y complejos son capaces de utilizar el código P (encriptado, por lo que es necesaria una clave correspondiente), así como las portadoras para un cálculo más preciso, según se explicó en un punto anterior.

Por último, y teniendo en cuenta que el sistema GPS mide las coordenadas (x, y, z) y el tiempo, y que existen diferentes precisiones en función de la tecnología que los receptores utilicen, encontramos una gran variedad de unidades receptoras, según estas se adapten para uno u otro uso principal. En líneas muy generales, los siguientes son algunos de los tipos principales en función de dicho uso.

- GPS para uso general. Unidades pequeñas y portátiles, de bajo coste, para actividades al aire libre, donde no se requiere una precisión elevada sino simplemente un conocimiento de la posición aproximada. Se emplean, por ejemplo, para recoger rutas en senderismo o navegación. Estas unidades, además de informar de la posición y ser capaces de almacenar esta, suelen disponer de capacidades de representación de mapas en pantalla, de forma que la información sobre la posición sea más útil para el usuario. Otros, como los navegadores GPS para coche, son capaces de calcular rutas óptimas, combinando la posición calculada con una cartografía de vías previamente incorporada al dispositivo. La figura 6.17a muestra un receptor GPS de uso general.
- GPS para la medición topográfica. Unidades de medio tamaño, generalmente con una antena independiente que se conecta a la unidad y que el propio operario carga a la espalda. La antena ga-

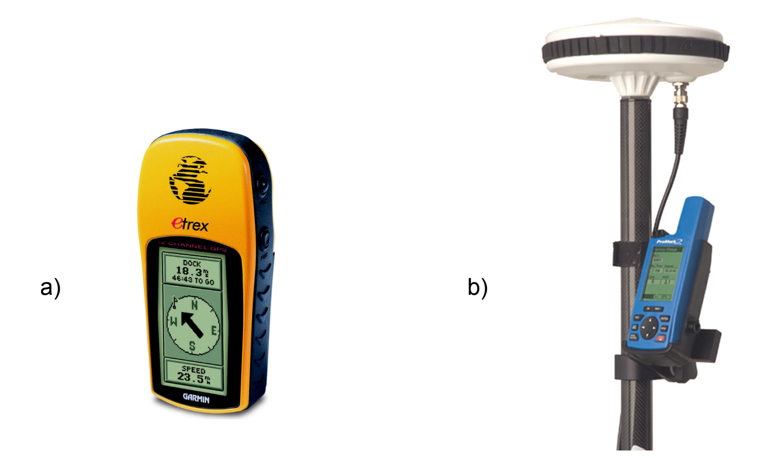


Figura 6.17: Receptor GPS de bajo coste para uso general (a) y receptor GPS de alta precisión con antena externa (b)

rantiza mayor precisión y una mejor localización de satélites en condiciones tales como zonas bajo arbolado. Están pensados para un uso profesional en levantamientos o replanteos, ofreciendo buena precisión en todas las coordenadas. En la figura 6.17b puede verse unos de estos receptores. Estos son los GPS de mayor interés para el uso dentro de un SIG, ya que ofrecen datos de campo precisos que cumplen con las necesidades que habitualmente se tienen en un proyecto SIG. Los datos recogidos por estas unidades pueden ser sencillamente incorporados a un ordenador, y en ocasiones la propia unidad dispone de aplicaciones propias, más allá de la mera visualización de cartografía asociada, como en el caso anterior.

- GPS para la medición del tiempo. Estos GPS no resultan de tanto interés para su uso en un SIG, ya que se encuentran fijos en un punto y no conceden importancia a la localización espacial, sino tan solo al tiempo. Se utilizan en estudios que requieran una medición muy precisa del tiempo, ya que la referencia temporal que ofrece el sistema GPS es muy precisa y estable.

6.6.3. Operaciones con la unidad GPS

La forma en que utilizamos el receptor GPS para recoger los datos que emplearemos posteriormente en el SIG puede ser muy variada en función del tipo de dato, la precisión necesaria o las características del propio receptor.

Los receptores de menor coste están generalmente pensados para ser de utilidad directamente en el campo, por ejemplo para localizar un punto concreto y conocer la dirección en la que hay que moverse para llegar hasta él, pero tienen también capacidad para recoger coordenadas. Estas capacidades son las que resultan de interés desde el punto de vista de un SIG, ya que las coordenadas recogidas serán después los datos que llevemos a este.

Por su parte, las unidades de mayor precisión están concebidas para tareas tales como levantamientos topográficos, donde la toma de datos es lo fundamental, pero también para otras tales como replanteos, donde se requiere situar un punto de coordenadas conocidas. Al igual que en el anterior, las actividades que pueden llevarse a cabo con estos GPS y que interesan desde el

punto de vista del SIG son aquellas que van a recoger coordenadas, pues son las que generan datos y convierten al GPS en una fuente de ellos.

Las capacidades de recogida de datos en una unidad GPS de bajo coste permiten almacenar puntos o trazados completos, encontrándose el operario inmóvil o bien en movimiento a lo largo de dicho trazado. Es habitual utilizar los vocablos ingleses de la terminología GPS para denotar los distintos elementos que pueden recogerse, conociéndose a un punto de interés aislado como *waypoint* y un trazado como *track*. Una serie ordenada de *waypoints* se conoce como *route* (ruta).

En el trabajo con el receptor GPS, el operario se puede detener en un punto cualquiera y memorizar las coordenadas del mismo, añadiendo así un *waypoint* a la lista de los ya almacenados. Para crear un trazado, se suele disponer de funcionalidades de recogida automática de puntos, de tal modo que el receptor memoriza estos a intervalos fijos de tiempo. El operario simplemente ha de desplazarse por el trazado y dejar que el receptor haga su trabajo mientras tanto. Dependiendo del tipo de dato que quiera obtenerse, la edición posterior en gabinete habrá de ser más o menos intensa.

Esta edición no está relacionada solo con la introducción de correcciones, sino con la interpretación de los distintos puntos recogidos. Por ejemplo, para registrar el trazado de una calle, el operario puede recorrerla, pero es probable que no lo haga de forma perfectamente rectilínea. El trabajo posterior con el conjunto de puntos debe resultar en la obtención de una línea recta a partir de estos, y ello requiere la interpretación de los datos disponibles.

Pese a que la precisión de estas unidades es limitada y no permiten técnicas avanzadas de corrección (tal precisión no es necesarias para las actividades tales como senderismo o navegación para las que han sido diseñados estos receptores), los GPS de uso cotidiano pueden ser una fuente de datos de primer orden para la recogida de datos. Un ejemplo significativo de ello es el proyecto OpenStreetMap, un proyecto colaborativo para crear mapas libres cuya principal fuente de datos son unidades GPS sencillas. Este proyecto es uno de los muchos que existen actualmente de este tipo, los cuales se engloban dentro de la idea de *Información Geográfica Voluntaria o Participativa*, sobre la que hablaremos algo más adelante en el apartado 6.7.

Para trabajos de mayor precisión tales como levantamientos topográficos, estos receptores no son, sin embargo, suficientes. El uso de receptores más precisos y de técnicas avanzadas es necesario para obtener precisiones mayores, que pueden ser incluso milimétricas.

Estos receptores pueden emplearse individualmente del mismo modo que se hace con un GPS de bajo coste, y registrar puntos de forma similar. La verdadera potencia, no obstante, se obtiene cuando se realizan mediciones con la ayuda de una o varias unidades adicionales, las cuales aportan valores de referencia que permiten aumentar la precisión.

Entre el receptor móvil y el de referencia se establece una *línea base*, y en el cálculo de la posición lo que se calcula es el vector (x, y, z) que une a ambas. Se trata pues, de una medición relativa, ya que expresa la posición del receptor móvil a partir de la del receptor de referencia. Puesto que la posición de este último se conoce con gran precisión y ese vector también se calcula con precisión, la posición buscada que se obtiene es altamente precisa.

La principal ventaja con respecto a métodos topográficos clásicos es que no es necesario que haya visibilidad entre los dos receptores. De esta forma, puede utilizarse una estación de referencia aunque no sea visible desde un punto cuyas

coordenadas queremos medir, y las líneas base pueden ser de mayor longitud.

Otras ventajas también destacables son el hecho de que puede obtenerse una productividad mucho mayor, ya que una única unidad de referencia puede ser utilizada por varias unidades móviles.

El número de técnicas existentes en la actualidad para realizar este tipo de mediciones (ya sea con uno o con varios receptores) es variada. El hecho de que se busquen mediciones precisas hace que se realicen mediciones utilizando la fase de la portadora, que como vimos implica una mayor necesidad de tiempo para registrar correctamente una posición. En función de las características de la línea base y los requerimientos concretos del trabajo, serán unas u otras las más adecuadas para cada caso.

La diferencia principal entre estas técnicas es el tiempo necesario para la recogida de un punto. En general, un mayor tiempo equivale a una mayor precisión. Entre las técnicas habituales, cabe citar las siguientes:

- **Estático**. En base a dos puntos de referencia (con una unidad GPS fija en cada uno de ellos), se calcula la posición de un tercero en un punto dado. Se trata del método más preciso, pero requiere tiempos de observación muy largos (superiores a una hora), lo que lo hace inadecuado para levantamientos o replanteos. Este tipo de procedimientos se emplean casi exclusivamente en trabajos geodésicos y las líneas base pueden ser de gran longitud.
- **Estático rápido**. Igual que el anterior, pero con tiempos menores, del orden de 5–10 minutos por punto medido.
- **Cinemático**. Los tiempos son aún menores que en el estático rápido, del orden del minuto. El fundamento de la técnica es distinto a los anteriores, ya que tras la inicialización el receptor móvil puede desplazarse con más velocidad y no es necesario que se detenga durante un periodo largo de tiempo en cada punto, pero ello exige que durante el desplazamiento tanto la unidad móvil como la fija de referencia mantengan la recepción de las señales, que han de ser de al menos cuatro satélites (preferiblemente cinco), y los mismos para ambas unidades. Si alguna de ellas pierde la conexión, se hace necesario repetir de nuevo el proceso de inicialización [53].
Existe una gran variedad de procedimientos de tipo cinemático, cuya filosofía es esencialmente la misma, pero bajo nombres distintos. Aunque pueden existir diferencias en los fundamentos teóricos, la forma de proceder es en muchos casos muy similar. Técnicas como *Stop & Go* o *pseudocinemático* pueden incluirse en este tipo de métodos. En general, estos y otros se engloban bajo la denominación de procedimientos cinemáticos, aunque sus características sean distintas en cada caso.
Muchos de estos procedimientos vienen definidos por el equipo a utilizar, y los tiempos de paradas en cada punto medido, así como otros aspectos, son recomendados por el propio fabricante. La forma más correcta de llevar a cabo una toma de datos en campo, en este caso, es seguir las indicaciones concretas del fabricante de para cada producto.
Un caso particular dentro de los métodos cinemáticos es el *cinemático en tiempo real* (RTK)[22], en el que, a diferencia de los anteriores, las correcciones necesarias se efectúan en tiempo real y no requieren postproceso. Se trata de la técnica más actual, y proporciona al operario mediciones exactas de su posición de forma instantánea, con las ventajas que ello conlleva. Las mediciones son más precisas, ya que el operario que las toma conoce el valor recogido en el mismo momento de hacer la medición, y puede de esa forma rea-

[22]Real Time Kinematic

lizar una comprobación en el acto. Información más detallada sobre esta técnica puede encontrarse en [54].

Para profundizar más al respecto, en [55] puede encontrarse información sobre la realización de levantamientos con GPS.

En base a los ejemplos anteriores, y para concluir esta parte, podemos dar una clasificación de las operaciones con un receptor GPS en función de tres criterios básicos: el número de unidades que se emplean simultáneamente, el movimiento (o ausencia de él) del receptor y el momento en el que se obtiene el dato ya listo para su utilización posterior.

Según el número de unidades, tenemos:

- **Absolutas**. Se tiene un único receptor y un único operario. La posición se calcula con la información de los satélites, sin apoyo de otra unidad adicional.
- **Relativas**. Se emplea una unidad adicional a modo de referencia. Las medidas se basan en la información de los satélites y la que aporta dicha unidad de referencia, y la posición se calcula en relación a esta en lugar de en términos absolutos. Estas operaciones alcanzan un grado de precisión mayor que las de tipo absoluto.

Atendiendo al movimiento del receptor encontramos:

- **Estáticas**.
- **Cinemáticas**.
- **Variantes intermedias**.

Por último, en función de la obtención de datos, distinguimos:

- **En tiempo real**. Las correcciones pertinentes se realizan en el acto, y el resultado que se visualiza en el receptor o se almacena en este ya ha sido filtrado y corregido.
- **Con necesidad de postproceso**. Las correcciones se realizan en gabinete posteriormente, con información que el receptor no posee o no es capaz de procesar de modo inmediato durante su utilización.

6.6.4. Integración de GPS y SIG

La utilidad de un GPS como fuente de datos para el trabajo en un SIG es innegable. Multitud de trabajos que requieren la toma de datos en campo y la medición de coordenadas pueden efectuarse ventajosamente con equipos GPS, y la información derivada de ese uso puede ser posteriormente incorporada a un SIG.

EL GPS puede emplearse como una fuente de datos estática (se utiliza como herramienta para la creación de una capa de información geográfica y esta después se emplea en el SIG de la forma habitual), o bien para la obtención de datos en tiempo real. Los SIG sobre dispositivos móviles (véase el apartado 25.2) pueden aprovechar los receptores GPS que estos dispositivos habitualmente incorporan, y alimentarse con los datos de dichos receptores en tiempo real.

Un caso particular de esto son los cada día más populares navegadores GPS. Estos dispositivos aúnan el receptor GPS y una aplicación de tipo SIG que presenta un visor y permite ejecutar un número reducido de procesos, en concreto los de cálculo de rutas óptimas entre dos puntos a través de una red de comunicación (apartado 20.4). Uno de los puntos (el de destino) es fijado por el usuario, mientras que el punto de origen es el punto actual en que se encuentra el dispositivo, que se obtiene a partir del GPS.

Como herramientas estáticas, el trabajo en campo con un GPS genera un conjunto de puntos o de trazados, que pueden fácilmente transferirse al ordenador para poder trabajar con ellos. Este trabajo puede realizarse dentro de un SIG, ya que, o bien este incluye la capacidad de importar los archivos generados por el GPS, o el software que

acompaña a dicho GPS incorpora herramientas para ayudar en la comunicación entre SIG y GPS.

Además de la información posicional que deriva del sistema GPS, los receptores GPS pueden incorporar elementos que permitan la entrada de la componente temática asociada a las distintas entidades, es decir, los atributos.

Las funcionalidades incorporadas en el receptor suelen ser sencillas, pero permiten que desde este se pueda llevar a cabo todo el proceso de creación de la capa que posteriormente se empleará en el SIG. El trabajo de campo incluye de este modo tanto el registro y creación de las entidades como la edición de las propiedades no espaciales de estos. Existe, igualmente, la posibilidad de completar la fase de introducción de atributos en el SIG, durante el trabajo en gabinete, lo cual en ocasiones resulta más sencillo y práctico.

El volumen de trabajo que se requiere una vez que los datos han sido recogidos dependerá también de las necesidades de precisión que se presenten y del tipo de trabajo en que se enmarque dicha recogida de datos. La realización de correcciones y la edición avanzada de los datos no puede en ocasiones realizarse dentro de un SIG, ya que este no dispone de las herramientas necesarias para un tratamiento avanzado de los datos del GPS. El SIG está preparado para trabajar con las coordenadas que salen del GPS, pero este puede almacenar más datos (datos «en bruto»), que pueden procesarse en gabinete para la obtención de dichas coordenadas de forma más precisa. Para realizar esta tarea es necesario software especializado, y las funcionalidades del SIG se emplearán posteriormente, cuando ya se hayan verificado los datos del GPS y elaborado las capas correspondientes.

Para el lector interesado, una referencia completa sobre el uso de GPS de cara a la integración de los datos en un SIG es [56]. En el ya mencionado apartado 25.2 veremos con detalle la tecnología de los SIG móviles, un ámbito en el que SIG y GPS se unen para conformar herramientas conjuntas.

6.7. Información Geográfica Voluntaria

Hemos mencionado ya que los dispositivos tales como receptores GPS de bajo coste pueden emplearse para recoger información geográfica y crear datos geográficos, y que cuando esto se une a los conceptos participativos de la denominada Web 2.0, surgen iniciativas de gran interés en las que el usuario de a pie, sin necesidad de una formación específica como cartógrafo, puede aportar sus datos para que otros los exploten posteriormente. Aunque no se trata de una fuente de datos como tal, y los elementos y dispositivos empleados ya los hemos visto a lo largo de este capítulo, el cambio que supone la inclusión de una filosofía acorde con las ideas de la Web 2.0 es tan notable que merece ser tratado por separado. No se trata de un cambio en la propia toma o preparación de datos, o de una tecnología nueva que se aplique a estos, sino de un cambio social y filosófico que redefine el propio concepto de la información geográfica en lo que a la creación del dato geográfico respecta, y cuyas consecuencias son ciertamente importantes, ya que abren el ámbito de la creación cartográfica a grupo nuevo de personas.

Se conoce como *Información Geográfica Voluntaria o Participativa* (en inglés Volunteered Geographical Information, VGI)[57] al uso de Internet para crear, gestionar y difundir información geográfica aportada voluntariamente por usuarios de la propia red. El conjunto de herramientas y técnicas que emplean esos usuarios para aportar su información conforma lo que se ha dado en llamar *neogeografía*. La comparación

entre proyectos de creación de VGI y la bien conocida Wikipedia, tal y como se comentó en otro punto anterior en este mismo capítulo, sirve perfectamente para ilustrar qué es lo que entendemos por VGI y neogeografía.

En el caso particular de esta última, la neogeografía ha supuesto un profundo cambio en algunas de las ideas básicas de la cartografía, modificando asimismo la concepción tradicional de la información geográfica, sus características o el papel que esta venía desempeñando en muchos ámbitos (o incluso dándole un papel en campos donde con anterioridad el uso de información geográfica era escaso). Algunas de las ideas principales sobre la neogeografía son las siguientes:

- Popularización y democratización. La producción cartográfica ha estado siempre en manos de gobiernos u organismos, y en muchas ocasiones fuertemente censurada debido a su elevado valor estratégico. Con la VGI, la creación de información geográfica se democratiza y se convierte en un proceso participativo libre y sin restricciones. Se invierte el esquema «hacia abajo» de producción y uso de información geográfica.
- Los ciudadanos se convierten en «sensores» y tienen mayor consciencia de su realidad geoespacial.
- Se elimina parte del «misticismo» de la producción de información geográfica

En parte, estas ideas son también comunes a otros fenómenos basados en la Web 2.0, ya que todas se fundamentan en una mayor democratización de la información, sea esta geográfica o no. También se comparten algunos de los problemas o críticas que otros ámbitos han recibido al adoptar esquemas de producción similares. Por ejemplo, la calidad de la información es puesta en entredicho al promover la participación de todo tipo de personas, con independencia de su perfil. En el caso de la información geográfica, con una producción tradicionalmente como hemos dicho limitada a profesionales muy especializados, esto es especialmente relevante. Con la proliferación de la VGI, se da voz y poder sobre la información geográfica a individuos en gran medida sin formación, que no obtienen un beneficio tangible obvio y no pueden aportar garantías de veracidad o autoridad alguna. Esto puede plantear dudas lógicas acerca de la conveniencia de usar esa información.

No debe olvidarse, no obstante, que la Web 2.0 también tiene sus mecanismos de regulación, y que en otros casos ya se ha demostrado que, para otros tipos de información, la calidad y rigor de esta no es inferior a la creada con esquemas más clásicos y menos abiertos.

Por otra parte, el hecho de que se use equipo de bajo coste y los usuarios no sean técnicos especializados no es necesariamente un problema. Un usuario sin formación no está capacitado para efectuar un levantamiento topográfico preciso, pero sí para situarse delante de la puerta de una tienda y marcar su posición, añadiendo esta a un proyecto que catalogue los comercios de la zona y su localización. Este tipo de información geográfica, de puntos de interés muchas veces no recogidos en cartografía más especializada, constituye una gran parte de la VGI, y las metodologías e instrumental con que se crea son más que suficientes para otorgarle validez y precisión adecuada al uso del que posteriormente va a ser objeto.

6.8. Sobre cartografía de elevaciones

La cartografía de elevaciones es probablemente la de mayor importancia de entre todas las que se emplean de forma habitual dentro de cualquier proyecto SIG. Su relevancia deriva del hecho fundamental de que la practica totalidad de procesos que se estudian en un SIG tienen algún tipo

de componente relacionada con el terreno y su relieve, y por tanto puede obtenerse amplia información sobre dichos procesos a partir de una capa con datos de elevación.

Dedicaremos en este libro un capítulo entero, el 16, al conjunto de operaciones de análisis basadas en el MDE, que van desde el simple cálculo de pendientes hasta la extracción de parámetros más complejos. Asimismo, gran número de otras formulaciones que veremos en la parte dedicada a procesos tienen su principal aplicación sobre datos de elevación, en particular los métodos de interpolación que veremos en el capítulo 14, y que nos permitirán crear cartografía de elevaciones en formato ráster. Este es, como veremos, el formato preferido para el análisis de la cartografía de elevaciones, ya que ofrece un mayor abanico de posibilidades frente a otros.

Aunque el formato ráster es el más indicado para llevar a cabo los análisis correspondientes, la cartografía de elevaciones puede crearse originalmente con muy diversas características. De igual modo, y debido también a la gran importancia de este tipo de capas, su origen puede ser muy variado, ya que son muchas las técnicas distintas que existen para su creación. Es de interés, por tanto, exponer en este capítulo sobre fuentes de datos algunas de las ideas principales relativas a la creación de capas de elevaciones, las características de estas o las ideas fundamentales que residen tras las metodologías más importantes. Posteriormente, esto nos ayudará a entender mejor las restantes formulaciones y conceptos relativos al manejo y análisis de este tipo de cartografía, abundantes en este libro como ya se ha dicho.

A modo de resumen, he aquí una lista de metodologías a partir de las cuales puede obtenerse cartografía de elevaciones, gran parte de las cuales han sido tratadas con detalle antes en este mismo capítulo.

- **GPS**. Como ya sabemos, un GPS toma datos no solo de la posición que ocupa en coordenadas x e y, sino también su elevación. La utilización de GPS permite obtener una nube de puntos de elevación, aunque si esta ha de cubrir un territorio amplio y con cierta precisión en las medidas, resulta poco adecuado el trabajar con esta tecnología, ya que requiere mucho tiempo. Es más adecuada para obtener levantamientos precisos de áreas más reducidas, donde se demuestra como una herramienta sumamente eficaz.
- **Digitalización de curvas de nivel**. En ocasiones la cartografía de elevaciones ya existe, aunque no en el formato adecuado para su empleo en un SIG. Ya conocemos los métodos de digitalización de entidades, tanto manuales como automáticos, y ya sea en pantalla o en equipo especializado, y mediante ellos podemos digitalizar las curvas de nivel, obteniendo una capa de líneas con la información altitudinal que contiene un mapa topográfico habitual.
- **Estereografía**. A partir de pares estereoscópicos, y con el concurso de una estación fotogramétrica digital pueden delinearse líneas o puntos de una elevación dada, digitalizando así la información altimétrica. El procedimiento es similar a la simple digitalización de curvas de nivel, solo que en este caso estas no están presentes explícitamente en las imágenes de partida, y se infieren a partir de la visualización tridimensional de las mismas.
- **Interferometría**. La interferometría es una técnica cuyos fundamentos son en cierta medida similares a los de la estereografía, pues se basan en la información recogida de un punto concreto desde dos puntos distintos. Si en el caso de emplear simples imágenes esto permitía crear una imagen tridimensional, en el caso de la interferometría el estudio de las diferencias de fases entre las ondas recibidas en dos puntos distintos permite el cálculo de distancias. Se tra-

ta, por tanto, de un proceso automatizado, que requiere menos intervención que en el caso de la restitución fotogramétrica.

Un uso muy habitual de esta técnica es con los denominados *Radares de Apertura Sintética*[23], utilizado por ejemplo en el caso de la misión SRTM, que reseñamos anteriormente como producto importante. La medición desde dos puntos puede hacerse con dos pasadas de satélite (caso por ejemplo del ERS) o bien en una sola si la plataforma dispone de dos receptores separados una cierta distancia (caso del SRTM). En [58] puede encontrarse una descripción detallada de este tipo de técnicas y las etapas que comprenden.

- **LiDAR**. La técnica más avanzada en la actualidad es el uso de aparatos de altimetría basados en láser, como el LiDAR, que ya hemos visto en este mismo capítulo. El LiDAR ofrece posibilidades muy interesantes tales como la obtención de MDE y MDS (Modelo Digital de Superficie) por separado.

  El resultado de un trabajo con LiDAR es una nube de puntos, normalmente en un número muy elevado debido a la precisión del instrumento, la cual puede emplearse para crear otro tipo de capas, tales como capas ráster. El nivel de postproceso que se requiere para la obtención final de una capa es mucho menor que con otras técnicas.

A la hora de plantear un proyecto SIG, debe elegirse entre estas fuentes, tanto si se desea adquirir la cartografía ya elaborada como si se desea crearla a partir de otras fuentes. La variedad de opciones existentes es grande, y cada una de ellas tiene sus características peculiares. Para saber más al respecto, algunas referencias donde puede encontrarse una comparación entre las metodologías anteriores son [59], [60] y [61].

[23]Synthetic Aperture Radar (SAR)

6.9. Formatos de archivo

Conocer las distintas fuentes de datos es importante para elaborar una base de datos geográfica que permita obtener los mejores resultados posibles, pero también lo es el conocer la forma en que esos datos pueden guardarse y transmitirse. Los datos geográficos se van a almacenar en archivos, y existen muchos formatos de archivo distintos para recoger un mismo conjunto de datos.

Estos archivos son la materialización de los modelos de almacenamiento que veíamos en el apartado 5.4, y su existencia obedece a distintas razones. Pueden haber sido definidos por alguna casa comercial para ser utilizados en su software, por un colectivo, o bien pueden ser estándares internacionales definidos para tratar de homogeneizar la forma en que se presentan los datos dentro de un determinado ámbito de trabajo.

Datos de una misma procedencia pueden presentarse de forma distinta si se emplean diferentes formatos de archivo. Las circunstancias por las cuales se opta por uno u otro formato pueden basarse únicamente en el hecho de que el software empleado soporte o no dicho formato, pero deberían fundamentarse en las propias características del formato y lo adecuadas que estas son para recoger la información con la que trabajamos.

La existencia de muchos formatos de archivo dificulta el trabajo con los datos en un SIG, principalmente porque ningún SIG implementa la capacidad de poder «leer» todos los formatos existentes. La interoperabilidad y la comunicación entre distintos SIG, o incluso entre un SIG y otras aplicaciones (bases de datos, aplicaciones para manejo de imágenes, aplicaciones CAD) no es completa, y el aprovechamiento de todos los datos disponibles dentro de un proyecto requiere

normalmente tiempo para la gestión adecuada de datos en formatos variados.

Un problema más serio, no obstante, es el desconocimiento por parte de los usuarios de las implicaciones que tiene el uso de uno u otro formato, ya que en ocasiones no permiten aprovechar de modo pleno los datos de que se dispone. Por ejemplo, dentro de un SIG es habitual emplear datos procedentes de CAD. Los datos en un CAD se almacenan en formatos de datos definidos por esas aplicaciones CAD, los cuales han sido definidos para satisfacer las necesidades del ámbito de trabajo en el que se han desarrollado (el diseño asistido por ordenador). Aunque los SIG pueden leer esos formatos de archivo y se encuentra información muy valiosa almacenada en ellos, no son ideales para el manejo de capas de datos SIG (en este caso, capas vectoriales), y es importante conocer este hecho.

La existencia de librerías que actúan a modo de interpretes facilita el desarrollo de aplicaciones SIG con capacidades de lectura y escritura en muchos formatos distintos, pero aún así se requiere un cierto grado de comprensión de estos por parte del usuario.

Debemos pensar asimismo que los formatos de archivo no solo se emplean en un proyecto SIG para los datos de entrada, sino también para almacenar los resultados que se generan a lo largo de ese proyecto. Estos datos serán utilizados en el propio SIG en otras ocasiones posteriores, o bien en otros programas. De este modo, tomamos datos que pueden provenir de aplicaciones y fuentes diversas, pero también «damos» datos a esas aplicaciones, por lo que la comunicación es bidireccional. Puesto que es a través de archivos como dicha comunicación se produce, y estos tienen que tener un formato dado, el conocimiento de estos formatos mejora tanto esa comunicación como la potencialidad de nuestros datos para todo tipo de uso, ya sea dentro o fuera de un SIG.

En esta sección no se pretende describir todos los formatos existentes, ya que son demasiados y ello no tendría sentido. Se describirán solo los más populares (que no siempre han de ser necesariamente los mejores) para que el lector obtenga un conocimiento general de *cómo* se van a presentar sus datos, y a través de estos formatos se describirán los principales enfoques existentes, que son los que realmente ha de conocer un usuario de SIG para saber discernir si un formato es o no adecuado para sus datos y las operaciones que quiere aplicar sobre ellos.

Junto con estos formatos de archivo, en el capítulo 24.7 se presentan los estándares de datos, que también se emplean para el intercambio y almacenamiento de datos SIG, y que presentan una relación estrecha con el contenido de esta sección. El capitulo 8, que veremos dentro de esta misma parte, también guarda relación con este apartado, pues estudia las diferentes formas en que los SIG han solucionado a lo largo del tiempo el acceso a los datos, incluyendo entre ellas el acceso directo a archivos.

6.9.1. Formatos para datos ráster

Los formatos de archivo para datos ráster son muy abundantes, existiendo numerosas alternativas con diferencias en ocasiones notables entre sí. Debido a que uno de los datos ráster más habituales en un SIG son las imágenes, a los formatos de datos específicos para datos ráster hay que sumar aquellos ya existentes para el almacenamiento de imágenes, que son de por sí muy variados. Estos formatos, adaptados a la naturaleza particular de las imágenes de un SIG, pueden emplearse para almacenar datos ráster y son de hecho de uso habitual en el ámbito de los Sistemas de Información Geográfica.

Formatos para imágenes

Las imágenes son un tipo de dato muy habitual en un SIG y se corresponden con el modelo de datos ráster. Por ello, los formatos de archivo empleados para el almacenamiento de imágenes digitales se emplean también para las imágenes particulares que utilizamos en un SIG (por ejemplo, fotografías aéreas o mapas escaneados, según vimos antes en este mismo capítulo), e incluso para otros datos ráster que no son imágenes como tales, como por ejemplo un Modelo Digital de Elevaciones.

Los formatos de archivo para imágenes son adecuados para recoger los colores de las imágenes, pero esto no es suficiente a la hora de almacenar otros valores (por ejemplo, valores decimales) o bien cuando son necesarios un número más elevado de bandas, como en el caso de imágenes hiperespectrales.

Una imagen en blanco y negro o en escala de grises contiene una banda. Una imagen en color contiene tres, ya que los colores se expresan como una terna de colores básicos: rojo, verde y azul. Este es el fundamento del modelo de color RGB, en el cual todo color es la combinación de distintas intensidades de los anteriores colores básicos. Las intensidades de cada banda (o las intensidades de la única banda en el caso de una imagen en escala de grises) se expresan habitualmente con valores entre 0 y 255, un rango que resulta insuficiente para el manejo de otras variables tales como las variables físicas que pueden emplearse en un SIG, ya que estas presentan valores continuos.

En estos casos, los formatos de imágenes no son adecuados en su forma original, y deben o bien adaptarse o bien emplearse formatos más específicos que tengan en cuenta el tipo particular de datos que se almacenan.

Otro problema es la presencia de celdas sin datos. La existencia de celdas sin datos es un hecho que no contemplan los formatos de imágenes. A estas celdas se les asigna un valor establecido por defecto, el cual ha de definirse en el propio archivo para que después sea reconocido por el SIG (para que sepa que, donde aparezca ese valor, realmente no existen datos), pero muchos formatos de imagen no puede almacenarlo. Una posible solución es la utilización de formatos que permitan transparencia. En estos, se puede especificar un color como transparente, que a efectos de su utilización en un SIG puede considerarse como indicación de la ausencia de datos. Estos formatos, no obstante, no son los más adecuados para datos SIG, y esta solución no resuelve por completo esta deficiencia.

Otra carencia de la mayoría de formatos de imágenes es que no pueden recoger la referencia geográfica de la imagen. Salvo que las imágenes sean utilizadas en un SIG, no hay necesidad de que estas contengan información tal como el tamaño de píxel (los metros que cada píxel representa en la realidad) o las coordenadas de la zona que recogen. Por ello, las definiciones de los formatos de imagen, al estar pensadas para recoger meras imágenes digitales (y no imágenes de satélite o aéreas destinadas a un análisis espacial), no tienen en cuenta estas necesidades.

Una forma habitual de resolver esto es acompañar cada fichero de imagen con un pequeño fichero de texto plano donde se contengan los datos geográficos correspondientes a la imagen. Este fichero se denomina *World File*, y contiene los siguientes valores:

- Tamaño de celda en la dirección Este–Oeste
- Ángulos de rotación del plano respecto a los ejes X e Y. Estos valores son siempre iguales a cero.
- Tamaño de celda en la dirección Norte–Sur, con signo negativo
- Coordenadas x e y del píxel superior izquierdo de la imagen.

Este *World File* tiene el mismo nombre que el archivo de imagen, y su extensión se forma con la primera y la última letra de la extensión de dicho archivo, y la letra `w`. Así, para un archivo `imagen.tif`, se tendrá un archivo `imagen.tfw`. Cuando el SIG abre la imagen, busca dicho fichero y, en caso de existir este, toma de él la información que necesita para poder incorporar la imagen al SIG de forma completa, de tal modo que sobre ella puedan llevarse a cabo análisis espaciales u operaciones como la digitalización en pantalla (heads–up) que hemos visto anteriormente.

Por último, un aspecto importante de los archivos de imagen es el tipo de compresión que utilizan. Las imágenes con las que se trabaja en un SIG pueden ser muy voluminosas, y para almacenarlas es necesaria gran cantidad de espacio (puede ser del orden de *gigabytes* para el caso de imágenes de alta resolución). Por esta razón, los formatos de imagen, especialmente los que han sido creados específicamente para imágenes SIG, incluyen algún método de compresión para disminuir el volumen del archivo.

En el apartado relativo a los modelos de almacenamiento vimos algunas ideas sobre compresión, presentando la codificación *run–length*. Esta es una estrategia para almacenar la información de forma que se minimice el tamaño de los datos necesarios, y en base a los datos recogidos puede recuperarse toda la imagen de forma exacta. Es decir, la utilización de estas formas de compresión no supone una degradación de la información contenida en la imagen, y nada de esta se pierde en el proceso. Podemos comprimir y descomprimir la imagen tantas veces como queramos, y el resultado siempre será el mismo, fiel a la imagen original. Un formato de archivo que cumple esto se dice que emplea un método de compresión *sin pérdidas*.

Por el contrario, existen otros métodos de compresión *con pérdidas*, en los cuales se pierde información y la imagen resultante, además de ocupar menos espacio, tiene una menor calidad y no es exactamente igual a la original, sino simplemente muy similar a esta. Los algoritmos de compresión con pérdidas toman de la imagen original la información más importante para después recrear esta, ignorando la menos relevante, que se pierde en aras de obtener un menor volumen de almacenamiento.

Siempre que sea posible, los formatos de compresión sin pérdidas deben preferirse frente a los que utilizan algoritmos de compresión con pérdidas. En función de las necesidades que se tenga con respecto a las imágenes a almacenar, debe elegirse el formato adecuado, considerando siempre la degradación que la compresión con pérdidas implica.

Algunos formatos de imagen que emplean compresión con pérdidas son altamente populares, ya que se emplean para tareas donde la reducción de tamaño de los ficheros es prioritaria, y este tipo de compresión ofrece una reducción en general mayor que la de los algoritmos sin pérdidas. Así, por ejemplo, las imágenes que se incorporan en paginas Web han de ser de pequeño tamaño para agilizar su carga, y ese tamaño resulta un factor decisivo, especialmente donde la velocidad de conexión es limitada. Para el trabajo con un SIG, no obstante, la calidad de la imagen es de mucho mayor importancia que su tamaño, y los formatos de compresión sin pérdidas responden mejor a las necesidades del almacenamiento de datos SIG.

En la imagen 6.18 puede verse el efecto de la utilización de compresión con pérdidas.

Formatos para datos SIG

Junto con los formatos de archivo para imágenes, los SIG ráster han desarrollado sus propios

a)  b)  c) 

Figura 6.18: Efectos de la utilización de algoritmos de compresión con pérdidas. a) Imagen original. b) Imagen almacenada mediante compresión con pérdidas. c) Imagen tras diez procesos de lectura y almacenamiento en un formato de archivo con compresión con pérdidas.

formatos para el almacenamiento de capas ráster en general, y en particular de aquellas que no representan imágenes, tales como capas de variables físicas.

Estos formatos están pensados para las características de estas capas, que habitualmente recogen valores decimales (a diferencia de los valores enteros de los Niveles Digitales de una imagen), y que no suelen contener más que una única banda.

Prácticamente cada SIG tiene su propio formato de archivo ráster, y otras aplicaciones que trabajan con este tipo de datos, tales como todas aquellas que usan por una u otra razón información de elevaciones, también disponen de sus formatos particulares. Muchos SIG pueden leer algunos de estos formatos junto con los suyos propios .

Principales formatos existentes

Dentro de la gran variedad de formatos existentes, he aquí una breve lista de los principales, los cuales suelen encontrarse con frecuencia a lo largo del desarrollo de un proyecto SIG habitual.

Dentro de los formatos para imágenes, cabe destacar los siguientes:

- **Tagged Image File Format (tif)**. Se trata de un formato complejo y altamente flexible, con muchas variantes distintas. Puede incorporar tanto compresión con pérdidas como sin pérdidas, en función del algoritmo que se utilice. Se utiliza habitualmente tanto en el ámbito del tratamiento de imágenes como en el ámbito SIG. En este último, permite también el almacenamiento de valores decimales, siendo apto para almacenar capas que no representen imágenes como tal. Es un formato habitualmente generado por los escáneres, con lo cual es frecuente su utilización al trabajar con cartografía escaneada, según vimos antes en este mismo capítulo. Existe una variante denominada GeoTIFF, que permite incorporar en el propio fichero la georreferencia de la imagen, haciendo innecesario el uso de un *World File* asociado.
- **Joint Photographic Experts Group (jpg o jpeg)**. Un formato muy popular para imágenes (todas las cámaras digitales lo utilizan), no es sin embargo adecuado para el trabajo con SIG. Incorpora compresión con pérdidas (el ejemplo de la figura 6.18 ha sido realizado utilizando este formato), y no es apto para almacenar capas ráster que no sean de tipo imagen.

Algunos formatos específicos para imágenes SIG tales como imágenes de satélite, son:

- **Enhanced Compression Wavelet (ecw)**. Está especialmente preparado para almacenar imágenes de gran tamaño, ya que las imágenes aéreas o de satélite en general tiene tamaños mayores que las imágenes de uso genérico para las que están pensados los formatos como TIFF o JPEG. En el uso de estas imágenes de gran tamaño en un SIG, es habitual que se quiera acceder a la imagen (por ejemplo para su visualización) solo en una parte determinada de la misma. Para optimizar este tipo de acceso, el formato soporta acceso sin necesidad de descomprimir la totalidad del archivo (descompresión selectiva). Se trata de un formato de compresión con pérdidas, y su grado de compresión es alto.
- **Multi–resolution Seamless Image Database (MrSID)** (sid). Al contrario que el anterior, que es un formato abierto, el formato MrSID es un formato cerrado, pero sus características son similares: alta compresión, preparado para imágenes de gran volumen y con posibilidad de descompresión selectiva.

Por último, entre los formatos para datos ráster (no imágenes) más comunes destacar el siguiente:

- **ArcInfo ASCII (asc)**. Un formato en texto plano ASCII[24]. Únicamente soporta una única banda, y permite almacenar el valor a considerar como valor de sin datos.

6.9.2. Formatos para datos vectoriales

Sin ser tan abundantes como los formatos para datos ráster, existe también un buen número de formatos de archivo para datos vectoriales. Al igual que en el caso ráster, estos formatos de archivo no derivan únicamente de los SIG, sino también de otras aplicaciones que utilizan capas de tipo vectorial, con particular importancia de las de diseño asistido por ordenador (CAD).

A la hora de definir las características de un formato de archivo para datos vectoriales, encontramos dos aspectos principales, a saber:

- Capacidad para recoger la topología de la capa
- Capacidad para recoger los atributos de las entidades.

En el primer aspecto, debemos considerar que existen SIG no topológicos, es decir, que no son capaces de manejar información sobre la topología de la capa, y por tanto no la necesitan. Los formatos de archivo de estos SIG no estarán por tanto pensados para trabajar con topología, y por ello no la almacenan.

Respecto a la capacidad para recoger los atributos de una capa, este aspecto afecta principalmente a los formatos propios de las aplicaciones CAD. En estas, la componente espacial es la que prima, no teniendo tanta relevancia la componente temática. Los puntos, líneas y polígonos con los que se trabaja en un CAD no tiene atributos asociados salvo aquellos relacionados con su propia representación tales como color, grosor o estilo. Existen formas de asociar una componente temática a esas entidades, pero estas son variadas y la interoperabilidad disminuye en caso de emplearlas, ya que no están soportadas con carácter general en los distintos SIG.

Por esta razón, estos formatos son aptos para introducir información dentro de un SIG o para exportarla a un CAD con objeto de utilizar capacidades de este que no se tengan en el SIG, pero como formatos de almacenamiento de datos dentro de un SIG no son los más adecuados, y debe optarse por otros más específicos para datos SIG.

[24] *American Standard Code for Information Interchange*. Un esquema de codificación de caracteres ampliamente utilizado.

Principales formatos existentes

Los formatos más extendidos para datos SIG vectoriales son los siguientes:

- **Shapefile (shp)**. Propuesto por la empresa ESRI, es el formato más utilizado en la actualidad, convertido en un estándar *de facto*. No soporta topología y se compone de diversos ficheros, cada uno de los cuales contiene distintos elementos del dato espacial (geometrías, atributos, índices espaciales, etc.)
- **Spatialite**. Una extensión espacial para la base de datos *SQLite*. Se trata de una base de datos, pero no tiene la arquitectura clásica de esta, con aplicación cliente y un servicio que provee los datos (lo veremos con más detalle en el capítulo 8), sino que toda ella se encuentra almacenada en un fichero que puede copiarse o eliminarse de la forma habitual. Basado en él, surge el formato GeoPackage, que cada día va cobrando más popularidad.
- **GeoJSON**. Un formato de texto plano basado en notación JSON[25], de uso extendido debido a su simplicidad. Existe una variante denominada TopoJSON, que permite el almacenamiento de topología.

6.10. Resumen

Los datos con los que trabajamos en un SIG pueden tener muy distintas procedencias. Distinguimos aquellos que provienen directamente de algún tipo de medida o del empleo directo de alguna instrumentación (fuentes de datos primarias), y otros que proceden de procesar un dato ya existente para adaptarlo a su uso en un SIG (fuentes de datos secundarias).

Una forma básica de crear datos espaciales digitales es la utilización de fuentes no digitales y su digitalización. Este proceso puede llevarse a cabo tanto de forma manual como automatizada, y puede dar como resultado tanto capas ráster como capas vectoriales.

La teledetección es una fuente de datos de gran importancia para los SIG. Dentro de ella se incluyen técnicas de muy diversa índole cuyos productos son muy distintos entre sí. El fundamento de la teledetección es la medición de las propiedades de los objetos realizada sin que medie contacto con estos. Para ello, se emplean sensores que pueden ir a bordo de aviones o montados sobre satélites, y que pueden ser de tipo pasivo o activo. El resultado del proceso de teledetección son imágenes con un número variable de bandas, aunque tecnologías como el radar o el LiDAR pueden emplearse para la generación de cartografía de elevaciones.

Dentro de las tecnologías que permiten la recogida de datos en campo, el GPS ha supuesto un cambio en la realización de este tipo de trabajos, y su integración en SIG es sencilla. Esto les ha convertido en una fuente de datos muy utilizada en un gran número de proyectos SIG.

Independientemente de su origen, los datos espaciales se almacenan en archivos cuyos formatos son a su vez muy variados. En este capítulo hemos visto algunos de los más habituales, así como los aspectos más importantes que los definen, y que han de tenerse en cuenta a la hora de trabajar con dichos formatos y elegir los más adecuados.

[25] JavaScript Object Notation

Capítulo 7

La calidad de los datos espaciales

7.1. Introducción

Puesto que los datos son la base de todo el trabajo que realizamos en un SIG, su calidad es vital para que ese trabajo tenga sentido y aporte unos resultados coherentes y útiles. Siendo la calidad el conjunto de propiedades y de características de un producto o servicio que le confieren su aptitud para satisfacer unas necesidades explícitas e implícitas [62], desde el punto de vista del SIG unos datos espaciales de calidad serán aquellos que puedan servir para alcanzar los objetivos de un proyecto concreto, dando sentido a este. En este aspecto, se debe considerar la disposición de los datos *per se*, aunque también las necesidades a las que pretendemos dar respuesta mediante los datos que utilizamos.

Por definición, ningún dato es perfecto. Todo dato que utilicemos va a contener errores, y estos pueden ser desde totalmente irrelevantes para el desarrollo de un proceso de análisis hasta de tal magnitud que desvirtúen por completo los resultados de dicho análisis. Es importante no solo contar con datos de calidad en los que estos errores sean mínimos, sino conocer el tipo de error que existe en nuestros datos y la magnitud de estos. Saber gestionar el error y ser consciente de las limitaciones de los datos de los que se dispone es importante para saber interpretar los resultados derivados del trabajo con dichos datos.

A lo largo de este capítulo veremos los aspectos más importantes que derivan de considerar el error como parte inevitable de nuestro trabajo con datos espaciales. Ello nos permitirá saber evaluar las capacidades de los datos para servir como punto de partida de nuestro trabajo, y a llevar este a cabo de la mejor manera posible, considerando que se trabaja simultáneamente con un conjunto de datos y con un error implícito asociado a estos.

[63] apunta las siguientes etapas para la modelación del error:

- Identificación de la fuente de error.
- Detección y medida del error.
- Modelación de la propagación del error.
- Propuestas de estrategias para la gestión y reducción del error.

Será sobre estas distintas fases sobre las que trataremos en las próximas secciones.

7.2. La importancia de la calidad de los datos

A pesar de su gran importancia, la calidad de los datos espaciales no ha sido una preocupación hasta hace relativamente poco tiempo. Los textos sobre Sistemas de Información Geográfica apenas trataban el tema en sus inicios [64, 65], y solo en la actualidad aparece una concienciación acerca de la importancia que la calidad de los datos

espaciales tiene sobre el desarrollo de cualquier trabajo basado en ellos.

Las razones por las que la calidad de los datos empieza a considerarse como un elemento de gran relevancia en el ámbito geográfico son principalmente dos [65]:

- Aparición de los SIG.
- Amplio crecimiento del volumen de datos espaciales disponibles, especialmente los derivados de satélites.

Estos dos factores, inevitablemente unidos, han favorecido que el volumen de trabajo sobre datos espaciales sea mayor y que además se use un número más elevado de datos distintos. Es lógico pensar que, a raíz de esto, haya surgido el interés por evaluar y tratar de forma rigurosa las condiciones en las que estos trabajos se están llevando a cabo.

La preocupación por la calidad de los datos es básica por el simple hecho de que datos de mala calidad generan inevitablemente resultados de mala calidad. Utilizar un dato de mala calidad es equivalente a utilizar un modelo equivocado. Si el modelo no es cierto, no importa la buena calidad de los datos, ya que los resultados que arrojará tampoco lo serán. Del mismo modo, un dato con un error superior al que puede resultar tolerable para una determinada tarea hace que la calidad de este sea insuficiente, y los resultados obtenidos carecen de valor.

A pesar de que la aparición de los SIG ha sido una de las razones principales para que se tenga en consideración la calidad de los datos y se especifique formalmente el modo de tratarla y gestionarla, los SIG en sí no disponen apenas de herramientas para asistir en estas tareas. Aunque la ciencia de la información geográfica ha avanzado mucho en ese sentido, y el conocimiento relativo a la calidad de los datos espaciales es mucho mayor, los SIG no han incorporado ese conocimiento y carecen de funcionalidades al respecto. Dicho de otro modo, existen las formulaciones y los elementos teóricos, pero estos aún no se han visto materializados (o lo han hecho de forma prácticamente anecdótica) en los SIG de uso habitual. Por esta razón, la mayoría de usuarios de SIG no tienen en cuenta rigurosa y formalmente la calidad de los datos a la hora de desarrollar su trabajo, quedando aún mucho por avanzar en este sentido.

Un elemento clave para el control de la calidad es la existencia de metadatos, que informan acerca de dichos datos sobre una serie de aspectos relativos a estos, entre ellos aquellos que afectan a la calidad. Los metadatos se tratan con gran profundidad dentro de este libro en el capítulo 9.

7.3. Conceptos y definiciones sobre calidad de datos

Antes de entrar en el estudio directo de la calidad de los datos espaciales y el estudio de los errores que pueden presentarse en un dato espacial, es necesario definir algunos conceptos básicos y alguna terminología al respecto.

El concepto básico es el *error*, que no es sino la discrepancia existente entre el valor real (puede ser un valor de posición, de un atributo, o cualquier otro), y el valor recogido en una capa. El error puede ser de dos tipos: *sistemático* o *aleatorio*.

Dos términos importantes en el estudio de la calidad son la *precisión* y *exactitud*. La precisión indica el nivel de detalle con el que se recoge la información. Una capa en la que las posiciones se han medido con 5 valores decimales es más precisa que una en la que se han medido con un único decimal.

Dependiendo del uso que se pretenda dar a una capa de datos geográficos, se requerirá una u otra precisión. Un trabajo geodésico requerirá medir la localización de un punto con precisión

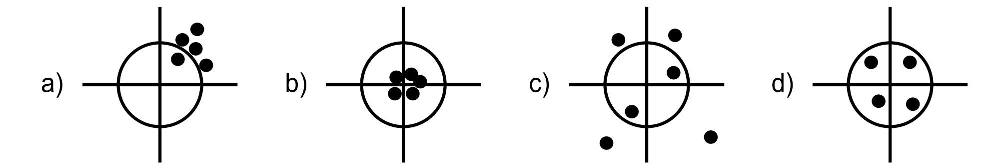


Figura 7.1: Diferencia entre precisión y exactitud (Tomado de [11]). En a) y b) la precisión es elevada, mientras que en c) y d) es baja. Por su parte, en a) y c) la exactitud es baja, siendo alta en b) y d).

milimétrica, mientras que para un muestreo para inventario forestal es suficiente localizar las parcelas correspondientes con una precisión mucho menor.

Por su parte, la exactitud nos indica el grado en que los valores estimados se asemejan al valor real.

La exactitud se calcula con el error sistemático, mientras que la precisión se calcula a partir del error aleatorio. Existe una relación directa entre precisión y exactitud, y en ocasiones se emplean ambos términos indistintamente. Si no existen errores sistemáticos (no existe un sesgo), la precisión y la exactitud son iguales.

Es posible, no obstante, que un dato sea muy preciso pero poco exacto, ya que las magnitudes de los distintos tipos de errores pueden ser muy distintas. Este hecho puede verse claramente en la figura 7.1.

Por último, un parámetro relativo al error es la *incertidumbre*. Habitualmente, el valor real es desconocido, por lo que el error no puede conocerse. La incertidumbre refleja la medida en que no podemos tener certeza de la validez de nuestros datos. La incertidumbre es un concepto más amplio que el error, y auna tres componentes [66]:

- **Error**
- **Vaguedad**. Aparece como consecuencia de definiciones pobres o incompletas, así como cuando los objetos que se modelizan en los datos no presentan límites bien definidos. Por ejemplo, en una capa de clases de vegetación, la transición entre una clase y otra se produce normalmente de forma gradual, por lo que establecer una frontera brusca es un hecho artificial que aumenta la incertidumbre, y el significado de asignar un punto en concreto a una clase dada es más vago cuanto más cerca de esa frontera nos encontramos.
- **Ambigüedad**. Cuando no existen definiciones inequívocas de los conceptos fundamentales, aparecen ambigüedades que añaden igualmente incertidumbre al dato creado en función de estos.

Tradicionalmente se ha trabajado con el error y no con el concepto de incertidumbre, pero conocer esta es igualmente importante a la hora de evaluar la calidad de los datos, y la modelización de la incertidumbre es una alternativa a la modelización del error.

7.4. Fuentes y tipos de errores

Cuando un dato espacial llega a nosotros para ser empleado en un SIG, ha pasado por una serie de etapas a lo largo de los cuales puede haber incorporado errores. Estudiando esas etapas por separado, encontramos las siguientes fuentes de error [67, 11]:

- **Errores de concepto y modelo**. Al recoger la información espacial utilizamos algún modelo de representación (ráster, vectorial), el cual siempre tiene alguna deficiencia. La realidad y las

tareas que pretendemos realizar con una capa de información espacial no se adaptan por completo a ninguno de los modelos de representación, y el hecho de optar por uno u otro conlleva la introducción de algún error, o condiciona para la aparición de unos u otros errores en las etapas posteriores.

- **Errores en las fuentes primarias**. El dato vectorial del que disponemos proviene originariamente de una fuente primaria, la cual puede contener errores. Si esta fuente contiene errores, estos aparecerán también en los datos que se deriven de este. Así, si digitalizamos en base a un mapa escaneado y la hoja original es errónea, también lo serán las capas que creemos en esa digitalización.
- **Errores en los procesos de creación de la capa**. Los procesos que realizamos para crear la capa pueden incorporar errores en el resultado. Por ejemplo, en el proceso de digitalización en base a ese mapa escaneado pueden aparecer errores por razones tales como un mal trabajo del operario, ya sea al digitalizar las entidades sobre una tableta o al teclear los valores de los atributos. Otros procesos, como pueden ser los de conversión entre los modelos ráster y vectorial, también pueden tener como consecuencia la aparición de errores. Los capítulos 14 y 18 tratan estos procesos de conversión, y se verán en su momento los posibles errores que pueden aparecer en cada caso y las razones por las que lo hacen. Igualmente, se verá cómo aplicar a esos procesos los elementos de medida del error que se desarrollarán más adelante en este capítulo.
- **Errores en los procesos de análisis**. Un dato espacial puede derivar de un proceso de análisis, y en él pueden aparecer errores debidos principalmente a dos razones: o bien la capa original objeto de análisis contiene de por sí errores, o bien el proceso no es por completo correcto.

  Veremos en el capítulo 16 cómo a partir de un MDE podemos calcular una capa con valores de pendiente, y cómo existen varios algoritmos distintos para realizar este cálculo. Ninguno de esos algoritmos es completamente preciso, y los valores calculados presentaran discrepancias de distinta magnitud con el valor real de pendiente, en función de diversos factores.

  Por su parte, el propio MDE también tiene sus errores, y estos se propagan a los resultados que derivamos de él, como veremos más adelante con detalle.

  En la parte dedicada al análisis veremos muchas operaciones que van a generar nuevos datos espaciales, y que pueden implicar la aparición de errores. Trataremos estos en su momento en la medida que ello pueda ser relevante para el manejo y utilización de esos datos derivados.

7.4.1. Las componentes de la calidad

La calidad de un dato espacial depende de muchos factores. Las características que dotan de dicha calidad al dato espacial son variadas, pues el dato espacial es en sí complejo, y cada una de estas características es susceptible de incorporar errores y por tanto de implicar una pérdida de calidad por ello. Los siguientes son algunos de los componentes principales de la calidad del dato espacial [65]:

- **Exactitud posicional**. Todo dato espacial tiene asociada una referencia geográfica. La precisión con la que se toma esta condiciona la calidad del dato.

  Esta precisión puede considerarse únicamente en los ejes x e y, o también en el eje z (elevación). Esta última, no obstante, puede considerarse como un atributo si se trabaja en un SIG bidimensional, y tratarse de la misma forma que cualquier otra variable de similar índole sin sig-

nificado espacial, tal como la temperatura en el punto (x, y) en cuestión.

- **Exactitud en los atributos**. Si la componente espacial puede tener errores, estos también pueden aparecer en la componente temática. Los valores asociados a una coordenada u objeto espacial pueden haber sido medidos con más o menos exactitud, o presentar valores incorrectos por muy diversas causas.
  Cuando el atributo en cuestión es de tipo categórico, puede existir un error de clasificación (se asocia la entidad espacial a una categoría errónea), mientras que en el caso de atributos no categóricos pueden sencillamente aparecer valores mayores o menores que los reales.
- **Consistencia lógica y coherencia topológica**. Los datos espaciales no son elementos independientes, sino que existen relaciones entre ellos. Un dato de calidad debe recoger fielmente estas relaciones, siendo la topología la encargada de reflejar este tipo de información. Por ello, debe existir una coherencia topológica en el dato espacial.
  Además de la coherencia de las relaciones, existe una coherencia implícita en todo atributo o valor recogido, de forma que resulte lógico. Estos atributos y valores han de ser coherentes con las escalas de medida o el tipo de valor que se espera, entre otros. Así, un valor de elevación no puede ser igual a «suelo calizo», ni un valor de temperatura expresado en Kelvin igual a -87.
- **Compleción**. El dato espacial no recoge todo lo que existe en una zona dada. Algunos elementos pueden no haberse recogido por cuestiones de escala (menores de un tamaño mínimo), pero también pueden incluirse o excluirse en función de otros criterios, en especial para el caso de mapas temáticos. Estos criterios deben conocerse para saber por qué un dato espacial contiene una serie de valores o elementos y no otros.
- **Calidad temporal**. Aunque los datos espaciales son «imágenes» estáticas de la realidad, el tiempo es importante en muchos sentidos, pues afecta directamente a su calidad. La realidad que representa un dato geográfico es una realidad que varía con el paso del tiempo, y por tanto este paso del tiempo puede degradar la calidad del dato espacial en mayor o menor medida.
- **Procedencia**. Un dato espacial puede provenir de una fuente más o menos fiable, o haber sido generado a través de uno o varios procesos, en cada uno de los cuales se puede haber introducido algún tipo de error. Conocer la procedencia de un dato y los procesos que se han empleado en su confección es necesario para poder evaluar su calidad.

Es importante recalcar que los errores que pueden incorporarse en estas componentes de la calidad pueden ser tanto de tipo cuantitativo como de tipo cualitativo, y que ello no está necesariamente ligado a la naturaleza de la componente o el tipo de variable a la que esta hace referencia. Así, un error en un atributo de tipo categórico supone un error cualitativo, pero un error posicional en la componente z (o de atributo de tipo continuo, si lo consideramos como tal) también puede dar lugar a un error cualitativo, como se muestra en la figura 7.2.

En la figura, que representa una porción de un Modelo Digital de Elevaciones y dos variantes alternativas con sendos errores de medición de la elevación, en el primer caso, y pese a que el error es mayor (hay mayor discrepancia entre el valor real y el recogido en el MDE), no varía la configuración del terreno. En la celda central encontramos una depresión, ya que en ella la elevación es menor que en las circundantes, y esto sigue ocurriendo así a pesar de existir ese error posicional. En el segundo caso (subfigura c), sin embargo, el error es menor en magnitud, pero al ser de signo contrario hace que la depresión

a)

11	11	11
11	10	11
11	11	11

b)

11	11	11
11	7	11
11	11	11

c)

11	11	11
11	12	11
11	11	11

Figura 7.2: a) MDE con valores reales. b) y c) Dos MDE con errores posicionales en z. En el caso c), el error no solo es cualitativo, sino también cuantitativo, ya que modifica la forma del terreno, pasando de ser una depresión a ser un pico.

se convierta en un pico, una configuración del terreno exactamente inversa. Si estudiamos las formas del terreno en ese punto (un análisis que arroja resultados cualitativos), obtendremos un valor erróneo.

Veremos más adelante que este tipo de errores son de gran importancia para muchos análisis, en particular para los relacionados con el comportamiento hidrológico del terreno, que estudiaremos en el capítulo 16.

La forma en que los distintos tipos de errores aparecen en una capa es diferente en función del modelo de representación empleado, ya que cada uno de estos modelos tiene sus propias debilidades, y las fuentes de datos de las que pueden proceder son asimismo distintas.

Así, los errores posicionales son más comunes en el caso de capas vectoriales, y una de las fuentes de error principal en este sentido son los procesos de digitalización, especialmente si son de tipo manual. Junto a los errores que vimos en el capítulo 6 (véase 6.5.5), existen otros que pueden aparecer al crear una capa vectorial, tales como los que se muestran en la figura 7.3 para el caso de digitalizar una línea.

Con independencia de la pericia y experiencia de un operador, resulta imposible que sea capaz de reproducir exactamente el objeto original y trazar con el cursor de la tableta digitalizadora o el ratón todos los detalles de este con absoluta fidelidad. Entre los errores que pueden aparecer encontramos falsos nudos (intersecciones de una línea consigo misma que no existen en realidad), puntos situados fuera del objeto, coincidencia imperfecta entre polígonos o mala referenciación de la hoja al situarla sobre la tableta (en el proceso de registro).

El problema principal en el caso de digitalizar líneas o polígonos (que pueden causar la aparición de mayor número de errores por su mayor complejidad) estriba en que aquello que se digitaliza es un conjunto infinito de puntos, y el proceso de digitalización solo puede recoger un número finito de ellos y después unirlos mediante segmentos rectilíneos.

La componente temática de una capa vectorial también puede adolecer de errores, que derivan a su vez tanto del proceso de introducción de los mismos como de los procesos de medición mediante los que se ha obtenido el valor concreto.

En el caso de capas ráster, la introducción de la componente temática no se realiza hacerse manualmente con el teclado, como sí puede sucede con las capas vectoriales. Ello no significa que las capas ráster no presenten errores en sus valores, pero el origen de estos es diferente. Un error habitual aparece en capas con información categórica que proceden de la clasificación de imágenes aéreas o de satélite. Los procesos que clasifican cada píxel de la imagen en función de sus Niveles Digitales (los cuales veremos en el capítulo 21) introducen frecuentemente errores,

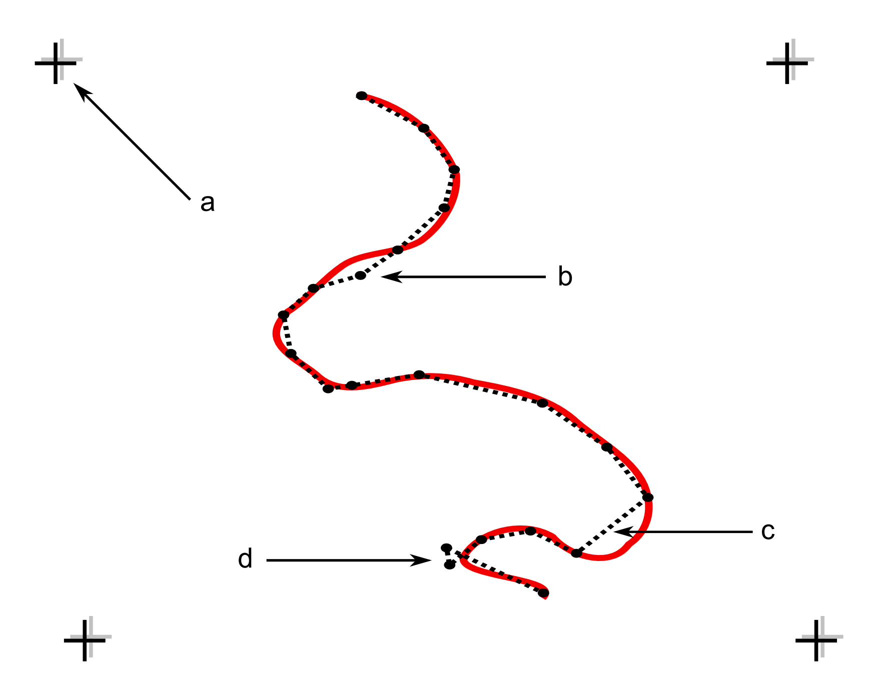


Figura 7.3: Algunos errores que aparecen en la digitalización de lineas. a) Registro inexacto, b) puntos mal situados, c) desplazamientos por vértices insuficientes, d) errores de registro.

y aparecen píxeles mal clasificados cuyo valor de clase no es correcto.

Los errores posicionales en las capas ráster se presentan de forma distinta a lo mostrado en la figura 7.3. Las entidades tales como líneas van a tener una representación errónea debido a la resolución de la capa ráster, que no va a permitir registrar con fidelidad su forma real. Por otra parte, la georreferenciación de una imagen incorpora asimismo errores, que son equivalentes al error de registro en la digitalización vectorial. Este error va a ser distinto según las zonas de la imagen, ya que la distorsión que implica la transformación realizada no supone un error constante. Veremos estas funciones con más detalle también en el capítulo 17, donde se tratan los dos principales errores que afectan a las imágenes: errores geométricos y errores radiométricos (básicamente, errores posicionales y errores en los Niveles Digitales).

Además de los errores de un único dato espacial (una capa de información), es importante considerar la forma en que los errores de distintos datos interactúan entre sí. En el trabajo con SIG es raro emplear una única capa, y lo más frecuente es trabajar con varias de ellas coordinadamente, cada una con sus respectivos errores. El modo en que esos errores se afectan entre sí puede condicionar la calidad de los resultados de forma similar a como los propios errores como tales lo hacen.

Como muestra la figura 7.4, dos errores sistemáticos de igual magnitud en sendas capas pueden tener efectos distintos sobre el resultado dependiendo de sus signos.

En la figura, tanto la capa de puntos como la de polígonos presentan un error sistemático. No obstante, un análisis que cuente el número de puntos dentro del polígono seguirá dando el mismo resultado en uno de los casos, ya que la

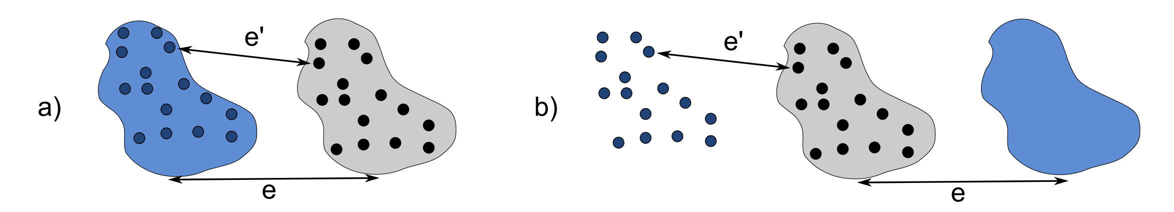


Figura 7.4: Un error e pueden tener distintas consecuencias según interactúen con los errores de otros datos espaciales (e'). En a) los errores casi se anulan, mientras que en b) se suman y dan lugar a un resultado erróneo. Los elementos en negro y gris indican la posición real.

forma de los errores de ambas capas hace que estos no afecten a este análisis, mientras que en el otro caso el resultado es completamente distinto del real.

7.5. Detección y medición de errores

Ahora que conocemos las fuentes y tipos de error, la evaluación y tratamiento de este empieza por su localización para saber a qué elementos del dato espacial afecta. Existen diversas metodologías para «inspeccionar» un dato espacial en busca de errores, que van desde métodos sencillos y obvios hasta avanzadas técnicas con base estadística para detectar patrones particulares o elementos «sospechosos» de contener algún error.

La forma más sencilla es la mera exploración visual. Algunos errores resultan obvios y una inspección sencilla permitirá localizarlos sin dificultad. Una coincidencia deficiente entre polígonos dejará un espacio en blanco que, si es de tamaño suficiente, puede ser localizado sencillamente en una exploración visual. De igual modo sucede con otro tipo de errores, en particular los errores de posición tales como los falsos nudos o la aparición de formas «ilógicas» (calles con ángulos muy bruscos, por ejemplo).

Es importante en este sentido que la representación del dato espacial sobre la que se realiza la exploración visual sea clara y adecuada, para revelar de la forma más notoria posible las posibles deficiencias de este. En este libro se dedica una parte entera a la visualización y representación de la información espacial y, al contrario de lo que pueda pensarse, esta no es solo de importancia para la generación de resultados al final de un flujo de trabajo, sino desde su mismo inicio. El análisis visual de los datos de partida, así como otros procesos de análisis, pueden beneficiarse de una representación correcta.

Existen errores que pueden detectarse visualmente, pero cuya detección (y corrección) puede automatizarse. Errores de este tipo son, por ejemplo, las conexiones imprecisas entre segmentos, que ya vimos en el capítulo 6. La función de *snapping* (ajuste por tolerancias), que se utiliza a la hora de digitalizar una capa vectorial, puede aplicarse *a posteriori*, una vez que la capa ya ha sido digitalizada. El SIG puede buscar esos enlaces imperfectos y convertirlos en enlaces correctos, resolviendo las uniones en las que exista una distancia entre vértices menor que una tolerancia preestablecida.

Como sabemos, hay SIG que son capaces de manejar topología y otros que no. También hay formatos de archivo que pueden almacenar topología y otros que no están pensados para ello. Por esta razón, los SIG topológicos trabajan a

menudo con datos sin topología, pero a partir de los cuales puede crearse esta, e implementan por ello las funciones para dicha creación de topología. Esta creación implica la corrección de errores topológicos que puedan existir en los datos originales, que no son relevantes en el caso de no trabajar con topología, y por ello pueden no haber sido detectados o eliminados. Errores como las antedichas falsas conexiones o los polígonos con adyacencia imperfecta, ambos se pueden corregir de forma automática, formando parte esas funciones de corrección de las rutinas de creación de topología.

Otros errores no pueden detectarse visualmente, en muchos casos porque los motivos del error no se representan y no aparecen en la visualización. Errores topológicos relativos a las estructuras de datos empleadas para recoger dicha topología entran en este grupo. En muchos casos, pueden no obstante corregirse de forma automática a través de operaciones de *filtrado* y *limpieza*, que se encargan de controlar la coherencia topológica del dato.

En el terreno de los atributos, la detección de errores puede llevarse a cabo empleando las técnicas estadísticas habituales. La detección de valores improbables (*outliers*) es uno de los procesos básicos. Estos *outliers* son observaciones dentro de un conjunto de datos que no parecen guardar consistencia con el resto del conjunto [68] y cuya detección puede llevarse a cabo de modo analítico o bien de modo visual, representando gráficamente los valores de los atributos. En general, las metodologías se fundamentan en comparar los valores con una distribución teórica y detectar la discordancia con esa distribución. Formas automatizadas de detectar *outliers* pueden encontrarse en [69].

Observaciones de este tipo, alejadas de las características generales del conjunto de datos, pueden derivar de medidas erróneas tales como las provocadas por un equipo de medición en mal estado, aunque también pueden representar valores correctos pero de carácter excepcional.

Si se combina la componente espacial con la componente temática encontramos otro tipo de valores inusuales, los denominados *outliers espaciales*. Estos se definen como observaciones que son discordantes con las observaciones realizadas en su vecindad[1] [70].

La diferencia entre un *outlier* en la componente temática y un *outlier* espacial es clara. Así, un valor de 10000 metros en elevación constituye siempre un valor excepcional, ya que va a encontrarse lejos de los valores medios recogidos, independientemente del lugar donde se hayan efectuado las mediciones. Un valor de 5000 metros puede constituir un *outlier* espacial en unas zonas (si tomamos medidas de elevación en, por ejemplo, Madrid, ya que será muy distinto del resto de elevaciones), pero puede ser un valor perfectamente lógico en otras zonas de estudio.

La detección de este tipo de valores puede realizarse, al igual que en el caso no espacial, de forma analítica o bien mediante exploración visual.

En base a lo anterior, existen una serie de procedimientos y metodologías para la detección de valores ilógicos en un conjunto de datos, los cuales se dividen de forma más genérica en dos grupos principales: unidimensionales y multidimensionales. Cuando en los multidimensionales la vecindad se define únicamente en función de la localización espacial y sin utilizar la componente temática, se tiene la detección de *outliers* espaciales. La figura 7.5 muestra un esquema de esta clasificación y las metodologías más habituales.

[1]Este hecho tiene relación con el concepto de *autocorrelación espacial*, que veremos en detalle en el capítulo 11, y que expresa la idea lógica de que las mediciones cercanas deben tener valores similares

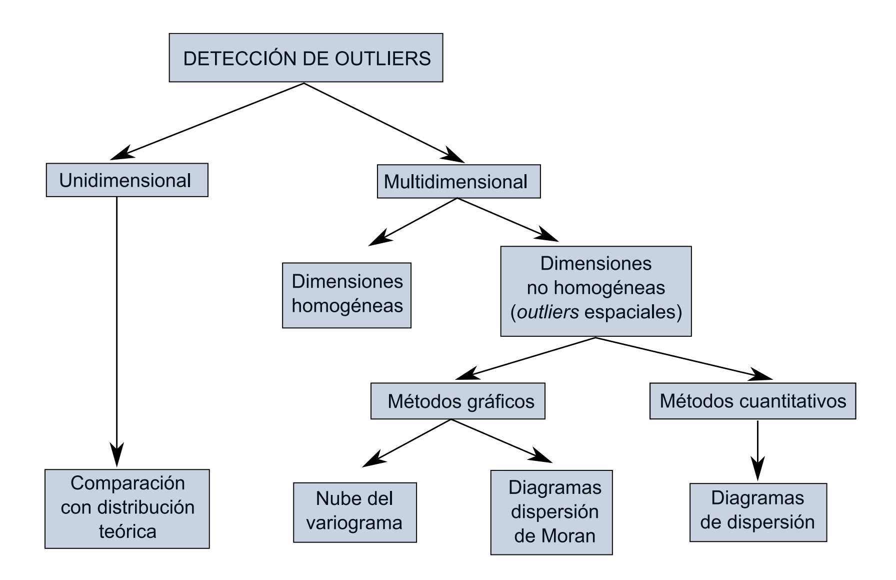


Figura 7.5: Clasificación de métodos para la detección de observaciones inconsistentes *(outliers)*

En [71] puede encontrarse más información al respecto.

Una vez localizado el error, este puede cuantificarse de diversas formas, según sea la naturaleza de la variable sobre la que se produce dicho error.

Los errores posicionales o los atributos no categóricos son variables de tipo cuantitativo. El *Error Medio Cuadrático* es la forma más habitual de medir esos errores. Su expresión es:

$$\mathrm{EMC} = \sqrt{\sum_{i=1}^{N} \frac{(y_i - y'_i)^2}{N}} \qquad (7.5.1)$$

donde N es el total de puntos en los que se comprueba el error, y el valor real, e y' el valor estimado. En esencia, se trata de una desviación típica, por lo cual se asume al emplear esta medida que los errores son aleatorios y se distribuyen normalmente.

Otras medidas utilizadas son el *Error Medio*, el *Error Medio Absoluto* o el *Error Máximo*.

Para valores cualitativos no puede aplicarse esta medida, y deben emplearse otros parámetros. La medida del número de valores que coinciden (elementos correctamente atribuidos) es una forma de determinar el error existente. El uso de la *matriz de confusión* es la forma más habitual de medir el error en la componente temática cuando esta es de tipo cualitativo. Veremos con más detalle su empleo y el de otras técnicas más complejas de similar propósito en el apartado 21.2.5.

7.6. Propagación de errores y modelación del error

El análisis de un dato espacial con errores va a dar un resultado que contiene a su vez errores, y existirá una relación directa entre los errores en el dato de partida y aquellos que aparecen en el dato resultante de su análisis. Este hecho se conoce como *propagación de errores*.

La propagación de errores puede ser muy variable en función del tipo de error que aparezca y la clase de análisis que se lleve a cabo. Errores de gran magnitud en el dato original pueden no tener apenas efecto en el resultado, mientras que

pequeños errores pueden causar grandes alteraciones en la calidad de este [72].

Una de las áreas en las que más se ha trabajado en el estudio de la propagación de errores es el trabajo con Modelos Digitales de Elevaciones. Como veremos en el capítulo 16, los MDE son un dato de primer orden, ya que resultan de utilidad en prácticamente cualquier tipo de proyecto SIG, y son muy numerosos los distintos parámetros que podemos derivar de ellos. Por esta razón, la propagación de errores es un asunto importante dentro del trabajo con un MDE, pues de él se van a obtener muchos datos nuevos, e interesa saber cómo la calidad de estos nuevos datos se va a ver afectada por la calidad del MDE de partida.

El error principal que se estudia en este tipo de análisis es el de los atributos, es decir, el de la elevación. Los datos empleados se basan en el modelo de representación ráster, ya que este es el más habitualmente empleado para los análisis de un MDE. No obstante, metodologías como la que veremos a continuación pueden aplicarse igualmente para la modelación de otros errores, tales como los errores posicionales en la digitalización de una capa vectorial.

La metodología más extendida para la modelación de errores es la basada en simulaciones de Monte Carlo. El fundamento de este método es considerar un dato espacial dado (un MDE para el caso de este ejemplo) como una de las posibles «versiones» de la realidad que pueden existir con una magnitud de error concreta. Evaluando el error existente en un dato espacial y su distribución, y realizando simulaciones estocásticas en base a este, pueden obtenerse otras de esas «versiones» de la realidad. Posteriormente, puede realizarse el análisis no sobre el MDE como tal, sino sobre todo ese conjunto de datos derivados del MDE y su distribución de error.

De este modo, se simula la presencia de error añadiendo ruido al MDE original, pero de una forma acorde con el propio error existente en el dato base. De las alternativas que se obtienen mediante estas simulaciones, ninguna de ellas tiene que ser necesariamente correcta y carente de errores [73] (lo más probable es que ninguna lo sea), pero el conjunto define un intervalo probable en el cual se situarán los valores reales. Se modela así la incertidumbre existente en el dato y la forma en que esta se propaga a los datos derivados.

Para una operación dada a aplicar sobre un MDE, la forma de proceder puede resumirse en los siguientes pasos [67]:

- Estudiar la distribución del error en el MDE en base a un juego de datos de referencia (generalmente un conjunto de puntos con mediciones precisas). Para modelizar el error no basta simplemente medir este con un parámetro como el error medio cuadrático, sino analizar su distribución y calcular parámetros estadísticos en base al conjunto de todos los errores medidos. Si se asume una distribución normal de los errores, la media y la desviación típica son necesarias para definir esa distribución.

  Al igual que sucede con los datos en sí, los errores tiene una dependencia espacial. Esto es, cerca de un valor que presenta un gran error, aparecerán otros también con errores notables, y cerca de valores donde el error es pequeño, no existirán puntos muy erróneos. La autocorrelación espacial, que veremos con detalle más adelante en este libro, se presenta tanto en los datos como en los errores.

  Por esta razón, la modelación del error requerirá conocer otros elementos adicionales para definir correctamente su distribución, tales como semivariogramas o correlogramas (estudiaremos estos en detalle en el capítulo 13, dedicado a la estadística espacial).
- Utilizando la distribución de los errores, se generan un número n de nuevos MDE. Para cada

uno de ellos, se genera una capa aleatoria de errores que se ajusta a la distribución definida, y esta se suma al MDE original. De este modo, en lugar de una posible versión de la realidad, se tienen n versiones.
La existencia de dependencia espacial puede añadirse en este paso si no se considera en el anterior, mediante el procesado de las capas de error y la aplicación de filtros sobre estas.
- Se aplica la operación sobre cada una de las n capas obtenidas.
- Se calculan parámetros estadísticos de los n resultados obtenidos, a partir de los cuales puede crearse un resultado único. Por ejemplo, la media de los n resultados obtenidos puede considerarse como valor resultante de la operación, en sustitución del que se obtendría aplicando esta únicamente al MDE original.

En la figura 7.6 se muestra un esquema gráfico de esta metodología.

Para ver con más claridad el efecto de este proceso, la figura 7.7 muestra respectivamente los resultados obtenidos a partir de un MDE, y la media de 20 y 50 simulaciones obtenidas según lo explicado anteriormente para el calculo de la curvatura horizontal (esté parámetro se explica en el capítulo 16).

Pese a su importancia, las herramientas para estos análisis no se implementan de forma habitual en los SIG, sino que deben llevarse a cabo utilizando funcionalidades individuales de análisis y programando los procesos repetitivos que son necesarios para calcular todas las capas derivadas empleadas. Por esta razón, es extraño que estos procesos se lleven a cabo en proyectos SIG de modo genérico. El usuario de SIG es consciente de los errores que presentan los datos espaciales con los que trabaja y las implicaciones de estos en lo que respecta a la calidad de datos, pero raramente desarrolla procesos de modelación de la incertidumbre, que quedan por el momento reservados para un ámbito más teórico que práctico.

7.7. Gestión de errores

Conocidos los tipos de errores fundamentales que encontramos en los datos espaciales y la manera de medir estos y su propagación, deben formularse estrategias para tratar de reducir el error, así como definir metodologías que permitan obtener resultados más precisos dentro de un proyecto SIG.

Estas estrategias dependen, como es lógico, del tipo de proyecto, sus objetivos, o el tipo de dato que se emplee para su desarrollo.

Podemos dividir estas estrategias en dos grupos fundamentales:

- Utilización de datos de partida más precisos. Deben establecerse parámetros de calidad referidos a los datos con los que se trabaja, que permitan tener garantía de que estos están en condiciones de dar respuestas correctas a las cuestiones que planteemos en base a ellos.
- Minimización de los errores a lo largo del desarrollo del trabajo. No todas las operaciones que realizamos en un SIG implican la introducción de errores en la misma medida. La propagación del error puede controlarse si estructuramos adecuadamente los pasos a realizar, situando al final de la cadena de procesos aquellos que sean más propensos a generar errores o sobre los que se tenga más incertidumbre en cuanto a la calidad de los resultados que arrojan.

Con independencia de la forma en que la gestión de errores se aborde, es importante que a la hora de trabajar con un SIG se tengan en cuenta ciertas ideas fundamentales con objeto de evitar la introducción de errores innecesarios. Algunas de estas ideas se enumeran seguidamente:

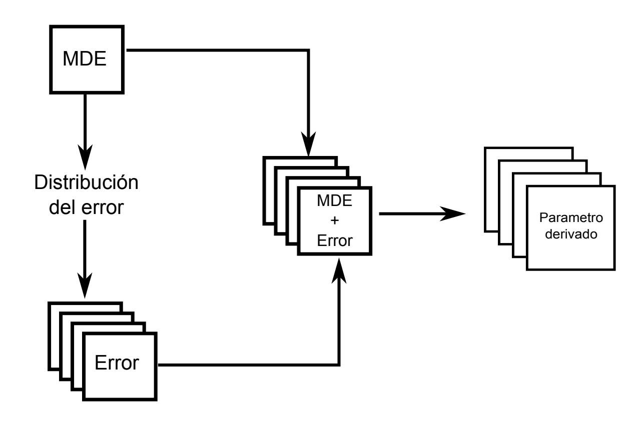


Figura 7.6: Esquema de la modelación de errores mediante simulaciones de Monte Carlo

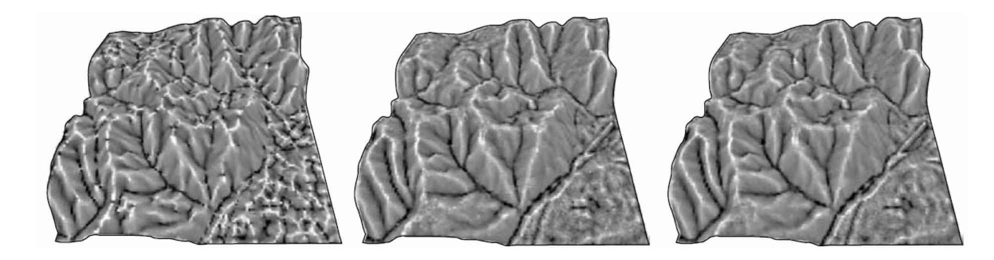


Figura 7.7: De izquierda a derecha, curvatura horizontal obtenida a partir del MDE original o como media de 20 y 50 simulaciones de Monte Carlo (tomado de [72])

- La utilización de capas de distintos orígenes y en distintos formatos favorece la aparición de errores y puede dar lugar a resultados de precisión insuficiente [74].
- La precisión disminuye a medida que lo hace la resolución espacial [75].
- La precisión de un resultado nunca sera superior a la del dato de entrada con peor precisión [76].
- Cuanto mayor es el número de capas empleadas para un análisis, mayores oportunidades existen de incorporar error a este e imprecisión a los resultados [76].

Es igualmente importante recalcar el hecho de que los datos digitales con los que trabajamos en un SIG no son *per se* mejores que los datos analógicos en cuanto a su precisión y su falta de errores. Si bien existen muchas ventajas asociadas a los datos digitales, tal y como vimos en el capítulo 6, la precisión no ha de ser necesariamente una de ellas, o al menos no como para poder asumir que su naturaleza digital implica que un dato es de calidad suficiente. En ocasiones, los usuarios de SIG pueden olvidar esto y trabajar bajo unas suposiciones incorrectas, introduciendo errores en sus resultados y no siendo conscientes de ello.

La importancia de los metadatos es grande en este sentido, ya que la cartografía impresa habitualmente contiene información acerca de su calidad y su precisión, pero al trabajar con una capa en un SIG, esa información la contienen los metadatos. Mientras que en un mapa impreso no podemos separar el mapa en sí de esa infor-

mación, en el contexto de capas de un SIG estas se encuentran formalmente separadas, hasta tal punto que la práctica más habitual es trabajar con capas sin metadatos o, de existir estos, no emplearse como parte importante de los propios datos.

7.8. Resumen

Pese a no haber sido una preocupación importante en los comienzos de los SIG, la calidad de los datos geográficos es hoy en día un aspecto clave para el trabajo con SIG. Las etapas fundamentales relativas a la calidad de los datos son la identificación de la fuente de error, su detección y medición, su modelación y, por último, la gestión de dicho error.

Las fuentes de error principales son las deficiencias de los datos originales, los errores conceptuales, los derivados de los procesos de digitalización y los introducidos en la realización de procesos con los datos. Estas fuentes introducen errores de posicionamiento, errores en los atributos asociados o de coherencia topológica, entre otros. Estas son algunas de las denominadas *componentes de la calidad*, entre las que también encontramos la procedencia de los datos o su validez temporalc.

Los errores aparecen de forma distinta en función de las características de los datos, en particular del modelo de representación elegido.

Detectar los errores puede realizarse de forma visual o bien de forma analítica, pudiendo automatizarse en este segundo caso. El error medio cuadrático es la medida más habitual del error en el caso de variables cuantitativas, mientras que la matriz de confusión es empleada para variables cualitativas.

Modelar el error y su propagación puede emplearse para conocer de forma más adecuada la validez de los resultados obtenidos a partir de un dato espacial. La realización de simulaciones condicionales mediante el método de Monte Carlo es la técnica más habitual para la modelación de errores.

Por último, es importante ser consciente de los errores que contienen los datos y de la posible aparición de estos a medida que realizamos tareas con ellos, con objeto de minimizar dicha aparición y limitar la presencia e influencia de los errores en los resultados finales.

Bases de datos

8.1. Introducción

Las bases de datos son un elemento fundamental en el entorno informático hoy en día y tienen aplicación en la práctica totalidad de campos. Concebidas con un propósito general, son de utilidad para toda disciplina o área de aplicación en la que exista una necesidad de gestionar datos, tanto más cuanto más voluminosos sean estos. Además, el trabajo con un SIG presenta una serie de características (uso múltiple de los datos, necesidad de acceso eficiente para análisis, necesidad de indexación, etc.) que hacen que sea recomendable el uso de bases de datos.

Pese a que, como veremos en este mismo capítulo, el uso de las bases de datos en el ámbito SIG no ha sido siempre el actual, hoy en día representan una parte clave para la gestión de los datos geográficos, en especial dentro del marco de proyectos de cierta envergadura. Aunque la realidad es que todavía se efectúa mucho trabajo SIG sin emplear bases de datos (y las aplicaciones SIG así lo permiten, no siendo estrictamente necesario disponer de una base de datos para almacenar la información), la naturaleza propia de los proyectos SIG y la progresiva implantación de los SIG a niveles más allá del uso personal traen ambas consigo un uso cada vez mayor de las bases de datos, y por tanto una mayor necesidad de conocer el funcionamiento de estas.

8.2. Fundamentos de bases de datos

Aunque las particularidades de los datos espaciales con los que trabajamos en un SIG han hecho necesarias modificaciones y adaptaciones sobre el esquema de trabajo de las bases de datos genéricas, en esencia los fundamentos de estas siguen constituyendo el elemento primordial sobre el que la arquitectura de gestión de datos espaciales se apoya. En esta sección, veremos de forma introductoria esos fundamentos de bases de datos genéricas, aplicables a cualquier otro ámbito además del de los SIG, para posteriormente poder tratar el caso particular de los datos espaciales.

8.2.1. ¿Qué es una base de datos?

Entendemos como *Base de Datos* un conjunto de datos estructurado y almacenado de forma sistemática con objeto de facilitar su posterior utilización. Una base de datos puede, por tanto, constituirse con cualquier tipo de datos, incluyendo los de tipo puramente espacial (geometrías, etc.) tales como los que se utilizan en un SIG, así como, por supuesto, datos numéricos y alfanuméricos como los que constituyen la componente temática de la información geoespacial. Los elementos clave de la base de datos son esa estructuración y sistematicidad, pues ambas son las responsables de las características que hacen

de la base de datos un enfoque superior a la hora de gestionar datos.

Podemos ver más claramente las implicaciones de utilizar una base de datos si recurrimos al ejemplo que vimos en el primer capítulo de este libro, relativo a la gestión forestal de un territorio. Para ello, consideremos que el número de usuarios del SIG y de los datos asociados no se limita únicamente al gestor forestal que ha de tomar decisiones o establecer planes de actuación, sino a muchos otros profesionales que puedan ejercer su trabajo en ese mismo área o puedan emplear total o parcialmente esos mismos datos.

Imaginemos, por ejemplo, el caso de un ingeniero encargado de planear la instalación de un tendido eléctrico a través de nuestra zona forestal de ejemplo. Sin duda, deberá emplear datos tales como Modelos Digitales de Elevaciones, capas de zonas protegidas o capas de arbolado para establecer el trazado óptimo y estimar costes de la línea, entre otras tareas. Si en una situación ideal este ingeniero estaría en comunicación con el gestor forestal y ambos compartirían sus conocimientos dentro de un equipo multidisciplinar, también en lo referente a los datos debería existir una comunicación igual que implique, ente otras cosas, un uso compartido y convenientemente coordinado de ellos. En otras palabras, los datos también tienen ese carácter multidisciplinar y deben dejar de verse como algo propio de un uso particular, para concebirse como un conjunto global del que se benefician diversos usuarios.

Establecer un uso compartido de los datos en una situación como la anterior no parece difícil, ya que simplemente se trata de dos profesionales que realizan tareas relacionadas y que, de un modo u otro, van a tener un contacto directo. El gestor forestal puede sencillamente dar una copia de sus datos al ingeniero y este podrá trabajar después con ellos de forma independiente. Aunque los datos con que trabajan son inicialmente los mismos, en realidad esta práctica da lugar a dos copias aisladas que constituyen dos universos distintos.

La situación real, sin embargo, es habitualmente mucho más compleja, y utilizar un esquema de colaboración como el anterior puede ser imposible, carecer por completo de sentido, o tener un buen número de consecuencias negativas. A medida que aumenta el número de usuarios, resulta menos recomendable que cada uno trabaje con sus propios datos y se los hagan llegar entre ellos a medida que los necesitan (una realidad que, desgraciadamente, se presenta con más frecuencia de lo recomendable). No debe olvidarse que un conjunto más amplio de usuarios que trabajan de esta forma y son ellos mismos quienes gestionan sus propios datos, implica directamente un número también más elevado de aplicaciones informáticas y de formatos de archivo, complicando enormemente el trabajo coordinado en cuanto el equipo tiene un tamaño medio.

Es probable además que existan usuarios dentro de una misma organización (por ejemplo, un organismo público) que aunque requieran para su trabajo datos similares, no tengan contacto alguno entre sí. Aunque los usuarios sean independientes, sus datos no lo han de ser necesariamente, y en una situación ideal deberían acudir a un repositorio único de datos del que cada cual tomaría lo necesario, en lugar de basar su trabajo en un conjunto de datos fragmentado y difícil de gestionar.

Pensemos en un dato que pueda ser de interés a varios usuarios, como por ejemplo una capa de vías de comunicación. A nuestro gestor forestal le será de interés para, por ejemplo, saber qué medios de acceso existen en caso de tener que hacer frente a un incendio. Lo más relevante de esas vías será su trazado, es decir su geometría, y tal vez el tipo de vía de que se trata, para poder conocer la velocidad a la que se pueden despla-

zar los medios de extinción. Otros usuarios, por su parte, pueden necesitar parámetros distintos como el volumen de tráfico medio de cada vía. Si todos ellos tienen una capa de vías con los parámetros asociados que necesitan para su trabajo, nos encontramos con una innecesaria redundancia de la componente espacial (las geometrías), y una dispersión de la componente temática, que resultaría más conveniente mantenerla agrupada.

Pensemos ahora que el gestor forestal detecta un error en el trazado de una de las vías y lo corrige. Esa corrección no estará disponible para los restantes usuarios, que pueden a su vez efectuar modificaciones similares que no redundarán en una mayor calidad de los datos con los que trabaja el gestor forestal, ya que, pese a utilizar datos similares, trabaja con su propio conjunto de datos. Incluso si en algún momento todos estos usuarios deciden poner en común sus datos y unirlos, esta operación puede ser muy compleja o incluso, como sucede frecuentemente, imposible de realizar. Por su parte, otros usuarios pueden añadir una nueva variable temática, como por ejemplo un índice de siniestralidad de la vía, el cual, si bien tal vez no resulte de utilidad inmediata para muchos usuarios, en un futuro sí pudiera serlo. Una vez más, estos nuevos datos no quedan a disposición del resto de usuarios, y en caso de serlo, no lo hacen en conjunto con datos similares, sino como un dato aislado de los restantes.

En definitiva, es complejo gestionar de forma adecuada los datos en el momento en que estos alcanzan un ámbito más allá de lo personal, y las prácticas más habituales basadas en una gestión «manual» de un conjunto de ficheros no son una opción adecuada. La solución para lograr esa necesaria gestión centralizada de los datos son las bases de datos y también, como veremos más adelante, los sistemas gestores de bases de datos, que representan la interfaz entre las bases de datos y los distintos usuarios.

8.2.2. ¿Por qué interesa usar una base de datos?

En base al ejemplo anterior, podemos analizar algo más sistemáticamente las ventajas de una base de datos frente a una gestión no organizada de los datos. Algunas ventajas que afectan directamente a los datos son las siguientes:

- **Mayor independencia**. Los datos son independientes de las aplicaciones que los usan, así como de los usuarios.
- **Mayor disponibilidad**. Se facilita el acceso a los datos desde contextos, aplicaciones y medios distintos, haciéndolos útiles para un mayor número de usuarios.
- **Mayor seguridad (protección de los datos)**. Por ejemplo, resulta más fácil replicar una base de datos para mantener una copia de seguridad que hacerlo con un conjunto de ficheros almacenados de forma no estructurada. Además, al estar centralizado el acceso a los datos, existe una verdadera sincronización de todo el trabajo que se haya podido hacer sobre estos (modificaciones), con lo que esa copia de seguridad servirá a todos los usuarios.
- **Menor redundancia**. Un mismo dato no se encuentra almacenado en múltiples ficheros o con múltiples esquemas distintos, sino en una única instancia en la base de datos. Esto redunda en menor volumen de datos y mayor rapidez de acceso.
- **Mayor eficiencia en la captura, codificación y entrada de datos**.

Esto tiene una consecuencia directa sobre los resultados que se obtienen de la explotación de la base de datos, presentándose al respecto ventajas como, por ejemplo:

- **Mayor coherencia**. La mayor calidad de los datos que se deriva de su mejor gestión deriva en mayor calidad de los resultados.

- **Mayor eficiencia**. Facilitando el acceso a los datos y haciendo más sencilla su explotación, la obtención de resultados es más eficiente.
- **Mayor valor informativo**. Resulta más sencillo extraer la información que los datos contienen, ya que uno de los cometidos de la base de datos es aumentar el valor de estos como fuente de información.

Por último, los usuarios de la base de datos también obtienen ventajas al trabajar con estas, entre los que cabe citar:

- **Mayor facilidad y sencillez de acceso**. El usuario de la base de datos se debe preocupar únicamente de *usar* los datos, disponiendo para ello de las herramientas adecuadas y de una estructura solida sobre la que apoyarse.
- **Facilidad para reutilización de datos**. Esto es, facilidad para compartir.

De forma resumida, puede decirse que la principal bondad de una base de datos es la centralización que supone de todos los datos con los que se trabaja en un contexto determinado, con las consecuencias que ello tiene para una mejor gestión, acceso o estructuración de estos.

8.2.3. Modelos de bases de datos

En función de la estructura utilizada para construir una base de datos, existen diversos modelos. El modelo de la base de datos define un paradigma de almacenamiento, estableciendo cómo se estructuran los datos y las relaciones entre estos. Las distintas operaciones sobre la base de datos (eliminación o sustitución de datos, lectura de datos, etc.) vienen condicionadas por esta estructura, y existen notables diferencias entre los principales modelos, cada uno de ellos con sus ventajas e inconvenientes particulares. Algunos de los más habituales son los siguientes:

- **Bases de datos jerárquicas**. Los datos se recogen mediante una estructura basada en nodos interconectados. Cada nodo tiene un único padre, y cero, uno o varios hijos. De este modo, se crea una estructura en forma de árbol invertido en el que todos sus nodos dependen en última instancia de uno denominado *raíz*. Aunque potente, el modelo jerárquico presenta algunas deficiencias, principalmente la escasa independencia de sus registros (el acceso a un registro —un nodo— implica que se ha de pasar por sus padres, restando flexibilidad a la navegación por la base de datos). Otra grave deficiencia de este modelo es la mala gestión de la redundancia de datos, ya que si un registro guarda relación con dos o más, debe almacenarse varias veces, al no estar permitido que el nodo correspondiente tenga varios padres. Esto tiene consecuencias no solo en el mayor volumen de datos que se almacena, sino también en la integridad y coherencia de los datos. Si se modifica una de las «copias» de ese registro en la base de datos, deben modificarse también las restantes, ya que, aunque no conectadas en la estructura de la base de datos, realmente representan una única realidad y debieran ser idénticas entre sí.
- **Bases de datos en red**. Con objeto de solucionar los problemas de redundancia de las bases de datos jerárquicas, surge el modelo en red. Este modelo permite la aparición de ciclos en la estructura de la base de datos (es decir, no ha de existir un único padre para cada nodo), lo cual permite una mayor eficacia en lo que a la redundancia de datos se refiere. Presenta, no obstante, otros problemas, siendo el más importante de ellos su gran complejidad, lo que hace difícil la administración de la base de datos.
- **Bases de datos relacionales**. Constituyen el modelo de bases de datos más utilizado en la actualidad. Solucionan los problemas asociados a las bases de datos jerárquicas y en red, utilizando para ello un esquema basado en tablas, que resulta a la vez sencillo de comprender y fácil

de utilizar para el análisis y la consulta de los datos. Las tablas contienen un número dado de *registros* (equivalentes a las filas en la tabla), así como *campos* (columnas), lo que da lugar a una correcta estructuración y un acceso eficiente.

- **Bases de datos orientadas a objetos**. Se trata de uno de los modelos más actuales, derivado directamente de los paradigmas de la programación orientada a objetos. El modelo extiende las capacidades de las bases de datos relacionales, de tal modo que estas pueden contener objetos, permitiendo así una integración más fácil con la propia arquitectura de los programas empleados para el manejo de la base de datos, en caso de que estos hayan sido desarrollados mediante programación orientada a objetos. Su popularidad crece de forma notable en ciertas áreas en las cuales resultan más ventajosas que el modelo relacional, siendo los SIG una de ellas.

La figura 8.1 muestra una comparación esquemática de los anteriores modelos de bases de datos.

8.2.4. Bases de datos relacionales

Aunque, como ya hemos visto, existen diversos tipos de bases de datos, las más utilizadas con diferencia en la actualidad son las relacionales, que han demostrado su idoneidad en la mayor parte de situaciones. Estas son también las que encontraremos en el ámbito SIG, y resulta por ello necesario añadir algunas nociones adicionales sobre ellas para la correcta comprensión no solo de este capítulo, sino también de otros posteriores que desarrollan temas relacionados.

El modelo relacional fue desarrollado en 1969 por Ted Codd y publicado un año después en un artículo ya clásico [78], y consiste básicamente en un conjunto de relaciones tabulares. Estas relaciones son tan importantes como los propios datos (las tablas, en este caso), y constituyen una idea central en el modelo relacional, de ahí su denominación. La característica principales que ha convertido a este modelo de base de datos en el más popular en la actualidad es su gran simplicidad, la cual indirectamente le dota de una gran potencia. Paralelamente, el modelo relacional se sustenta en unos fundamentos matemáticos sólidos y sus ideas pueden expresarse mediante conceptos de la teoría de conjuntos, lo que posibilita un análisis formal del mismo.

Además de las denominaciones habituales de *tabla*, *fila* y *columna*, existe una terminología específica empleada al referirse a las bases de datos relacionales. Así, en el modelo relacional los datos se organizan en tablas bidimensionales, cada una de ellas con información relativa a un determinada *entidad*. La tabla en sí se conoce como *relación*, ya que recoge la relación existente entre sus elementos, y constituye así el eje central del modelo relacional. Dentro de la tabla, los datos están organizados a su vez en filas y columnas. Las columnas representan los distintos *atributos* asociados a la entidad, mientras que las filas conforman los distintos *registros*. Una fila se forma con un conjunto de n atributos, constituyendo una *tupla*.

El *esquema de la relación* está formado por los nombres de los atributos y un *dominio* asociado a estos, que delimita el rango de valores posibles para cada atributo. El dominio especifica el tipo de dato a contener en cada columna. Por ejemplo, si se recoge un nombre el atributo será de tipo alfanumérico, mientras que si el atributo es un conteo deberá ser de tipo entero. Además de los tipos habituales (fechas, cadenas de texto, valores

[1]Entiéndase el adjetivo *real* aquí en su sentido matemático, es decir, un número n tal que $n \in \mathbb{R}$. Puede emplearse también la denominación menos formal de *número decimal* o bien *valor de coma flotante*, esta última más común en el ámbito informático y referida a la forma de almacenamiento de este tipo de valores.

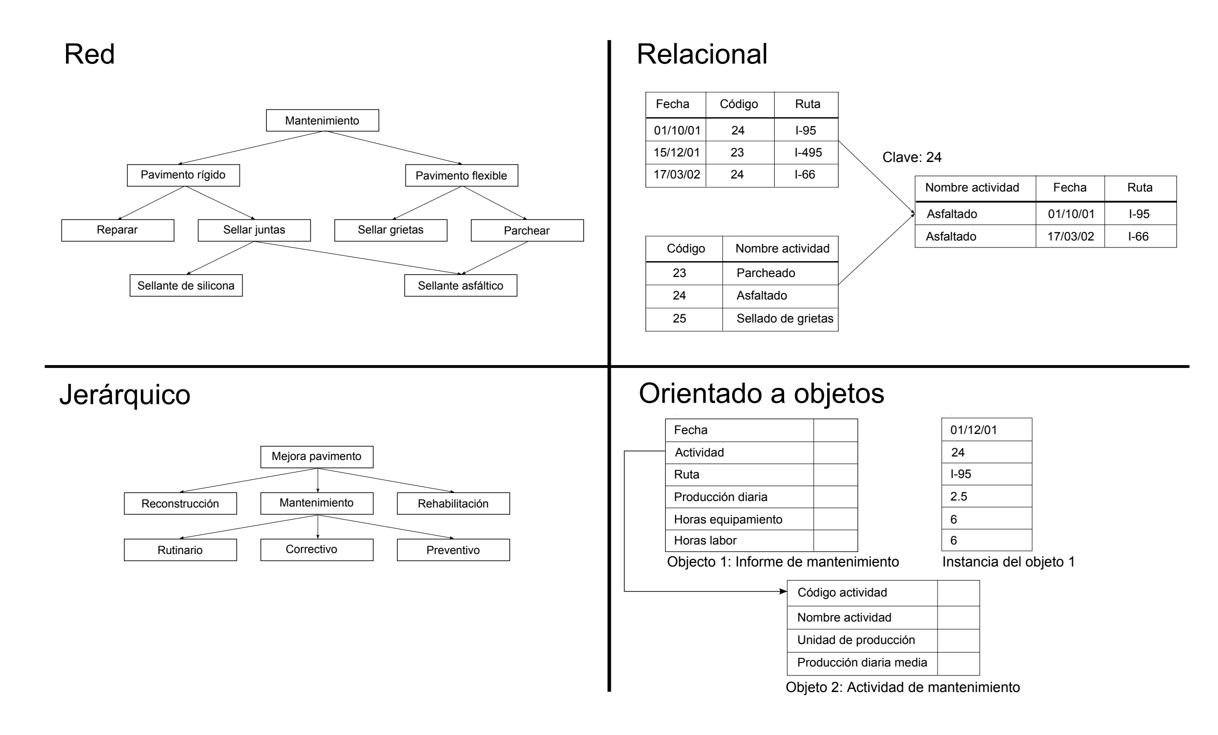


Figura 8.1: Comparación entre algunos modelos de base de datos más frecuentes (adaptado de [77]).

reales[1], valores enteros, etc.) pueden emplearse en ciertas bases de datos valores más complejos. Esto es de especial interés en el caso de los SIG, ya que permite utilizar geometrías como un tipo de datos más, con la utilidad que esto tiene a la hora de almacenar datos espaciales. El esquema de la relación se recoge en la primera fila de la tabla, conocida como *cabecera*. El número de filas de la tabla sin contar la cabecera (es decir, el número de tuplas) se conoce como *cardinalidad*.

Las relaciones son, por tanto, un conjunto de tuplas asociadas a un esquema. En una relación, tanto el orden de las filas como el de las columnas son irrelevantes (exceptuando la cabecera, que no es un tupla como tal, sino que define el esquema como hemos visto), pero es importante que cada atributo sea del tipo correspondiente a la columna a la que pertenece. Es decir, que sea coherente con el esquema.

En la figura 8.2 puede verse un esquema de los elementos fundamentales del modelo relacional.

Una forma abreviada de definir las relaciones que forman parte de una base de datos es mediante su nombre y su esquema expresado como una lista de los atributos que lo constituyen. Por ejemplo, podemos definir una relación denominada `PERSONAS` como

```
PERSONAS(DNI, Nombre, Altura, Edad,
         Ciudad)
```

Una base de datos contiene normalmente más de una tabla, ya que suelen ser muchos los tipos de datos a almacenar y resulta conveniente dividirlos en distintas tablas. Además de las relaciones que la tabla en sí implica, es necesario definir relaciones entre las distintas tablas, y para ello se emplean los denominados atributos *clave*. Un

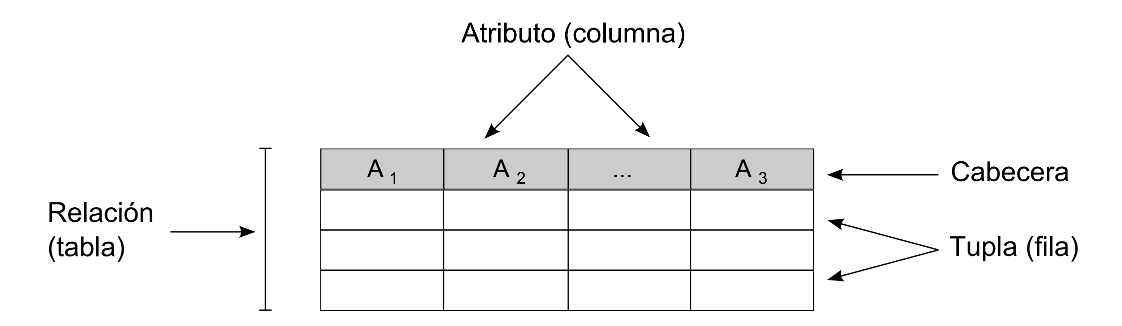


Figura 8.2: Elementos del modelo relacional.

atributo clave es aquel que tiene valor único para cada tupla, pudiendo servir para representar a esta plenamente. Por ejemplo, en una tabla con nombres de personas e información adicional sobre ellas según el esquema anterior, los nombres no pueden ser la clave primaria, ya que puede haber dos personas con un mismo nombre. El número de su Documento Nacional de Identidad, sin embargo, sí que puede servir como atributo clave. Además de su unicidad, una clave debe ser invariable, identificando la misma tupla a lo largo del tiempo. Un esquema de relación puede contener varios atributos clave, que se conocen como *claves candidatas*. Normalmente, de estas se elige una como representante principal de las tuplas, y se conoce como *clave primaria*

Por convención, las claves se escriben subrayadas al definir el esquema de la tabla, de tal modo que el de la tabla `PERSONAS` quedaría de la siguiente forma:

```
PERSONAS(DNI, Nombre, Altura, Edad,
          Ciudad)
```

Si no existe ningún atributo que cumpla los requisitos para ser utilizado como clave, este puede incorporarse al esquema de la relación, añadiendo por ejemplo un nuevo atributo con un código arbitrario. Un ejemplo de esto lo podemos encontrar en el cuadro 8.1, donde se incorpora un atributo que hace la función de clave a una tabla con información sobre personas pero que no contiene el DNI de estas entre esa información y, por tanto, carece de un atributo adecuado para servir de clave.

En la definición de clave cabe también la presencia de claves compuestas, es decir, formadas por varios atributos cuya combinación es única para cada tupla. No obstante, la utilización de claves simples es preferible generalmente, ya que simplifica gran parte de las operaciones en las que la presencia de una clave es necesaria.

Cuando trabajamos con datos espaciales, es habitual emplear la componente espacial como clave, ya que esta suele ser única. En el caso de almacenar información sobre ciudades, con los nombres sucede de forma similar a lo visto para el caso de personas, ya que existen ciudades con el mismo nombre en distintos lugares. La localización de estas, sin embargo, es única, ya que no puede haber dos ciudades simultáneamente en el mismo lugar.

El empleo de estas claves permite relacionar tablas entre sí, siempre que estas compartan algún atributo común. Por ejemplo, pensemos en una base de datos que contenga la tabla anterior y junto a esta otra tabla con el siguiente esquema:

```
CIUDADES(Nombre, Población,
         Superficie)
```

a)

DNI	Nombre	Altura	Edad	Ciudad
50234561	Juan Gómez	1,85	35	Madrid
13254673	Edurne Montero	1,60	30	Toledo
46576290	Luis Urrutia	1,75	46	Madrid
38941882	Juan Gómez	1, 71	55	Valencia

b)

ID	Nombre	Altura	Edad	Ciudad
001	Juan Gómez	1,85	35	Madrid
002	Edurne Montero	1,60	30	Toledo
003	Luis Urrutia	1,75	46	Madrid
004	Juan Gomez	1, 71	55	Valencia

Cuadro 8.1: Adición de un campo para crear una clave. La tabla a) contiene un atributo único (DNI). La tabla b) no contiene un atributo único entre sus datos, pero se añade el campo ID con un código arbitrario que puede ser empleado como clave. El nombre en este caso no sirve como atributo único, ya que hay dos personas en la tabla con el mismo nombre.

Es decir, la base de datos contiene información sobre personas y sobre ciudades.

Es sencillo ver que puede vincularse una tabla a la otra a través del atributo que contiene el nombre de la ciudad. Nótese que este atributo no tiene el mismo nombre en ambas tablas, y que, mientras que en una de ellas representa la clave primaria[2], en la otra no puede serlo pues existen nombres de ciudades repetidos. Pese a ello, este atributo nos permite establecer una relación entre las tablas[3], que podríamos denominar «nacido en». A cada tupla de la primera tabla, que representa a una persona dada, podemos vincularla con una de la segunda tabla, que representa una ciudad en particular, ya que toda persona ha nacido en una ciudad y gracias al atributo `CIUDAD` podemos saber exactamente cuál es dicha ciudad.

Las interrelaciones entre tablas pueden ser de distintos tipos en función del número de elementos distintos que se vinculan de cada tabla. En nuestra relación «vive en», una persona puede vivir en una única ciudad, mientras que una ciudad puede tener muchas personas viviendo en ella. Es decir, cada tupla de la tabla `PERSONAS`

[2]Pese a que se ha comentado que el nombre de la ciudad puede no ser adecuado como clave, en este caso sí puede serlo debido a las pocas filas que contiene la tabla, por lo que, en aras de la simplicidad, lo utilizaremos asumiendo que no van a existir en la tabla dos ciudades con el mismo nombre.

[3]Nótese que estamos empleando aquí el término *relación* para referirnos al vínculo entre tablas, pero que este término también se emplea para referirse a las propias tablas, lo cual puede dar lugar a confusiones. Para evitarlo, emplearemos el término *tabla* para referirnos a estas, y mediante los términos *relación* o *interrelación* haremos a partir de ahora únicamente referencia a esos vínculos que permiten enlazar varias de dichas tablas.

se relaciona con una única de la tabla CIUDADES, y cada tupla de esta última se relaciona con una o varias de la primera. Este tipo de relación se conoce como de *uno a muchos*.

Existen otros dos tipos de relaciones además de esta: las denominadas de *uno a uno* y las de *muchos a muchos*. Un ejemplo de relación de uno a uno podrían ser «casado con», que estableceríamos entre la tabla PERSONAS y ella misma (las dos tablas implicadas no han de ser necesariamente distintas). Cada persona puede estar casada únicamente con otra, por lo que la relación es de uno con uno, relacionándose una tupla con tan solo otra distinta, y no con varias.

Es importante reseñar que en algunas relaciones como «nacido en» todos los elementos de una o de las dos tablas se encuentran vinculados de algún modo a través de la relación, mientras que en otros no es así necesariamente. Así, todas las personas han nacido en alguna ciudad, y estarán relacionadas con la correspondiente tupla en la tabla CIUDADES, pero no todas las personas están necesariamente casadas.

Un ejemplo de relación *muchos a muchos* la podemos plantear si contamos en nuestra base de datos con, por ejemplo, una tabla con empresas, entre cuya información se incluya una lista de las ciudades en las que cada empresa tiene sede. Una empresa puede tener sedes en distintas ciudades, y una ciudad puede acoger a varias empresas, con lo que tanto ciudades como empresas pueden estar vinculadas a más de una tupla en la otra tabla.

8.2.5. Sistemas gestores de bases de datos

Junto con las bases de datos, el elemento fundamental para el aprovechamiento de estas son los *Sistemas Gestores de Bases de Datos* (SGDB o DBMS, del inglés *DataBase Management System*). Estos sistemas representan un elemento intermedio entre los propios datos y los programas que van a hacer uso de ellos, facilitando las operaciones a realizar sobre aquellos. En nuestro caso, son el componente que permite unir el SIG con la base de datos en la que se almacenan los datos espaciales con los que este va a trabajar.

Un SGBD es una pieza de software compleja, ya que las situaciones a las que debe responder son diversas y en muchas ocasiones con requerimientos elevados, por ejemplo en lo que a eficiencia y volumen de datos respecta. Piénsese que una base de datos actual puede tener millones de registros y ser utilizada simultáneamente por miles de usuarios, que a su vez pueden utilizar diversos programas, no todos ellos del mismo tipo. Por ejemplo, una base de datos que contenga números de teléfono, nombres de usuarios, direcciones y coordenadas asociadas a cada línea telefónica, puede ser empleada desde un SIG para crear un mapa que muestre la densidad de usuarios o también desde una aplicación que genere un listín telefónico, o bien desde una aplicación en una página Web que permita localizar el número de teléfono de una persona concreta. Cada una de estas aplicaciones realiza un trabajo distinto, pero todas ellas utilizan la misma base de datos. El SGBD debe proporcionar a todos ellos la metodología adecuada para extraer del conjunto de datos completo cuanto sea necesario en cada caso.

Además, el SGBD es la herramienta utilizada no solo por quienes aprovechan los datos, sino también por aquellos que se han de encargar de la propia gestión y mantenimiento de la base de datos. Administrar una base de datos puede suponer una tarea altamente compleja, por lo que el SGBD debe proveer los útiles necesarios para llevar a cabo ese mantenimiento.

Para ser de verdadera utilidad y responder a todas las necesidades que pueden plantearse

en relación con la base de datos, un SGBD debe perseguir los siguientes objetivos:

- **Acceso transparente a los datos**. La base de datos ha de poder accederse de forma transparente, sin que sea necesario para el usuario del SGBD preocuparse por aspectos internos relativos a la estructura de esta u otras características. Esto significa que, por ejemplo, si queremos recuperar un registro de la base de datos, debemos poder hacerlo sin necesidad de saber si dicha base de datos está almacenada en un único archivo o varios, o si el registro que pretendemos recuperar está almacenado a su vez de uno u otro modo. Así, el SGBD debe crear una abstracción de los datos que haga el trabajo con estos más sencillo, ocultando aspectos que no sean relevantes para dicho trabajo.
Procedimientos como las consultas que veremos en el capítulo 12 se realizan a través del SGBD, que es quien se encarga de interpretar dichas consultas, aplicarlas sobre la base de datos y devolver el resultado correspondiente. El SIG no accede a los datos, sino que se comunica con el SGBD y deja en manos de este el proceso de consulta en sí.
- **Protección de los datos**. Si la base de datos almacena información sensible, el SGBD debe controlar el acceso a esta, restringiendo el acceso cuando corresponda (por ejemplo, estableciendo distintos permisos de acceso para distintos tipos de usuarios) e implementando los mecanismos de protección necesarios.
- **Eficiencia**. Acceder a los datos no es suficiente en la mayoría de los casos, sino que se requiere un acceso eficiente. El SGBD debe ser capaz de gestionar de forma fluida grandes volúmenes de datos o de operaciones (por ejemplo, muchos usuarios accediendo simultáneamente), de modo que dé una respuesta rápida a las peticiones de los usuarios de la base de datos.
- **Gestión de transacciones**. Las operaciones sobre la base de datos tales como la adición o borrado de un registro se realizan mediante transacciones. Una transacción es un conjunto de operaciones realizadas por un usuario sobre la base de datos como una única unidad de trabajo, de forma indivisible. El SGBD ha de encargarse de gestionarlas de manera eficiente y segura para que todos los usuarios de la base de datos puedan hacer su trabajo de forma transparente. Aspectos como el acceso concurrente a la base de datos (varias transacciones simultaneas) resultan especialmente importantes, y en su buena gestión se pone gran esfuerzo en el diseño de los SGBD.
Se denomina *transaccional* al SGBD capaz de garantizar la integridad de los datos, no permitiendo que las transacciones puedan quedar en un estado intermedio. Esto implica la capacidad de poder volver a un estado anterior en caso de que por cualquier causa (error en el sistema, fallo eléctrico, etc) no haya podido completarse la transacción.

La figura 8.3 esquematiza el papel que el SGBD juega en el manejo y empleo de los datos. Tanto los distintos usuarios (en el caso de nuestro supuesto de gestión forestal pueden ser desde el gestor forestal al cartógrafo encargado de actualizar los limites de las unidades inventariables) como el administrador de la base de datos acceden a esta a través del SGBD. No existe acceso directo a la base de datos.

El SGBD tendrá unas u otras características en función del modelo de base de datos subyacente, ya que debe adaptarse a las características de este para ofrecer las funcionalidades correspondientes en el nivel de usuario.

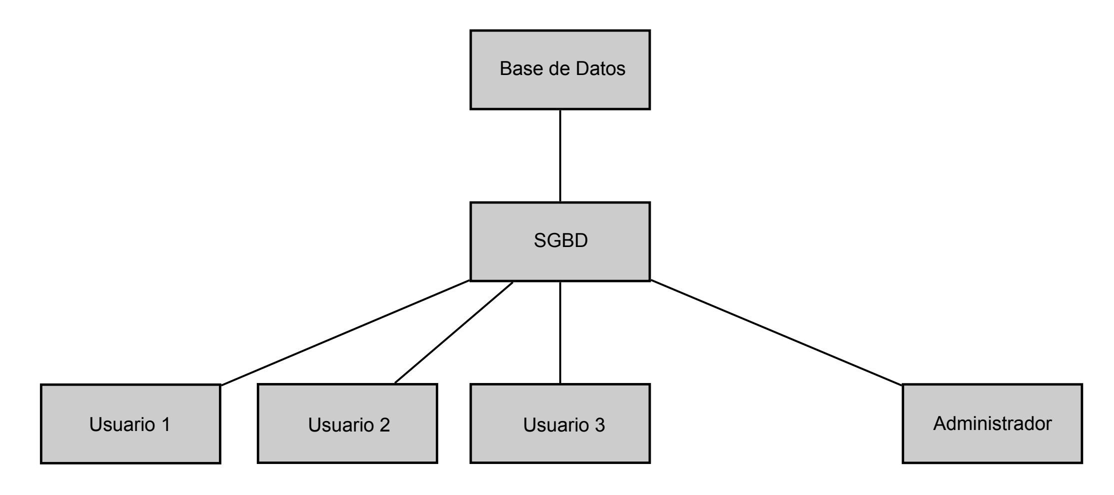


Figura 8.3: Representación esquemática del papel de un Sistema Gestor de Base de Datos.

8.2.6. Diseño y creación de una base de datos

Una vez se toma la decisión de emplear una base de datos, el siguiente paso es el diseño y creación de esta. El diseño implica la definición de la estructura que va a tener la base de datos, que se deberá realizar teniendo en cuenta principalmente el tipo de datos que van a almacenarse y el modelo de base de datos elegido. El diseño debe adecuarse al uso previsto de la base de datos, de tal modo que acomode los datos de la mejor forma posible para cumplir los objetivos enunciados anteriormente en este mismo capítulo. Para ello debe conocerse la naturaleza de los datos que van a almacenarse (no necesariamente datos de los que se dispone en el momento de la creación, sino los que se espera pasen a formar parte de la base de datos a lo largo de su ciclo de vida), así como la de los algoritmos y procesos que van a emplearse sobre ellos.

Posteriormente al diseño, debe procederse a la implementación de la base de datos, esto es, a la creación propiamente dicha, incorporando los datos según los esquemas escogidos en la fase de diseño. Por último, y una vez creada la base de datos, debe procurarse un mantenimiento para que esté continuamente en condiciones de ser utilizada.

Más concretamente, pueden distinguirse las siguientes fases en el proceso global de desarrollo de una base de datos:

- **Diseño lógico**. Independiente del SGBD empleado, es un diseño conceptual que pretende modelizar el contenido de la base de datos.
- **Diseño físico**. Es la adaptación del diseño conceptual a las particularidades del SGBD escogido.
- **Implementación**. Introducción de los datos en la base de datos.
- **Mantenimiento**. Monitorización de la actividad sobre la base de datos.

La primera fase en el diseño de una base de datos implica un análisis de los datos que se van a recoger. Como resultado de ese análisis debe surgir un modelo conceptual que exprese la estructura de la información, siendo dicha estructura susceptible de ser empleada como esquema de partida para la base de datos en cuestión. El modelo conceptual ha de definir básicamente los tipos de datos a tratar y las relaciones existentes entre ellos, elementos que serán luego expresados

en términos del modelo de base de datos elegido (relacional, orientado a objetos, etc.) una vez se pase a la fase de diseño físico.

El modelo conceptual debe estructurar la información de forma que el usuario de la base de datos comprenda de forma sencilla el contenido y forma de esta. Por tanto, debe desarrollarse teniendo presentes las necesidades de los usuarios y el hecho de que estos no necesariamente han de ser especialistas en bases de datos, sino especialistas en los propios datos en sí. Por otra parte, el modelo debe intentar capturar del mejor modo posible la realidad que se pretende modelizar, por lo que el conjunto de tipos de datos y relaciones debe elaborarse de modo similar a dicha realidad para recoger toda la complejidad del sistema. Y, por supuesto, el modelo debe poder ser implementado posteriormente y utilizado en conjunto con el SGBD escogido, ya que de otro modo no presenta utilidad práctica.

Existen diversas metodologías para desarrollar un modelo conceptual. Una de las más extendidas por su sencillez y potencia es la del *modelo entidad–relación* (abreviadamente, modelo E-R).

Denominamos *entidad* a un objeto o concepto del mundo real acerca del cual se recoge información, y que puede diferenciarse de otros objetos, incluso si son de su misma clase (un ordenador, por ejemplo, es un objeto, y puede diferenciarse de otros ordenadores, incluso si son de idénticas características, ya que no son todos el mismo objeto y ese en particular tendrá alguna propiedad distinta, como puede ser el número de serie). La entidad puede tener sentido físico o bien ser una idea abstracta, como un tipo de deporte, una clase de música o una palabra.

Una entidad se describe mediante una serie de características o atributos, que son las que definen su naturaleza y sus propiedades. Una colección de entidades es un conjunto de entidades distintas (que representan a objetos distintos), las cuales comparten unos atributos comunes. Por ejemplo, un conjunto de ordenadores de los cuales se conocen los atributos *modelo*, *marca* y *procesador*.

Por su parte, una *relación* expresa la dependencia existente entre entidades y permite la asociación de estas. No resulta difícil ver que estos conceptos —entidad, atributos y relación— guardan un notable paralelismo con las ideas del modelo relacional que ya conocemos. Así, y aunque no resulte por completo inmediato, es sencillo traducir un modelo entidad-relación (conceptual) a un modelo relacional, que constituye ya un modelo aplicado a un tipo particular de base de datos. Por ello, el modelo E-R es una herramienta potente para el diseño lógico de la base de datos, especialmente si esta utiliza el modelo relacional.

Para desarrollar el diseño conceptual de una base de datos siguiendo el modelo E-R, estos son los pasos principales:

- Partimos de una descripción textual del problema o sistema que queremos recoger. Esta descripción contiene los requisitos necesarios y ha de formular la pregunta a la que queremos que la base de datos dé respuesta. Para nuestro ejemplo con datos sobre personas y ciudades, el problema podríamos formularlo como «¿qué personas han nacido en cada ciudad?»[4].
- Se toman los verbos y los sustantivos de la descripción textual. Los sustantivos son posibles entidades o atributos, mientras que los verbos son posibles relaciones. En nuestro caso, «perso-

[4]Nótese que, aunque por meras razones didácticas hemos presentado en el capítulo las tablas correspondientes a personas y ciudades antes de abordar lo relativo al modelos E-R y el diseño de la base de datos, este modelo E-R es previo a la implementación de la base de datos, y las tablas correspondientes al modelo relacional son solo una implementación práctica de dicho esquema, en este caso según los requisitos de una base de datos que utiliza dicho modelo relacional

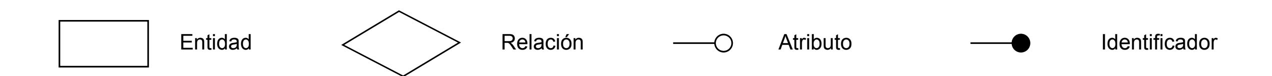


Figura 8.4: Simbología empleada en el modelo entidad–relación.

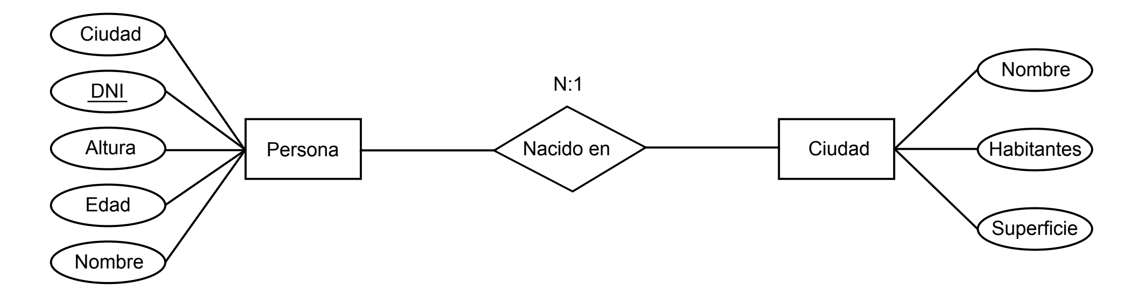


Figura 8.5: Ejemplo de diagrama E-R.

na» y «ciudad» serán entidades y «nacido en» una relación.

- Se analizan las frases y determina la cardinalidad de las relaciones y otros detalles.

El modelo así creado se expresa mediante un diagrama en el que las entidades se representan como cajas rectangulares, las relaciones mediante rombos y los atributos en círculos o elipses, todos ellos con sus correspondientes nombres en el interior. Cuando un atributo es un identificador, se representa con su nombre subrayado, del mismo modo que en la definición de esquemas que ya vimos anteriormente (Figura 8.4).

Si el número de atributos es elevado o el diagrama es complejo por existir gran cantidad de tablas e interrelaciones, pueden omitirse los atributos para una mayor legibilidad, describiéndose en un documento adicional.

Como ejemplo de lo anterior, la información sobre personas y ciudades que venimos manejando, así como la relación «nacido en» existente entre ambas, se expresarían según el modelo entidad-relación con un diagrama tal como el mostrado en la figura 8.5.

El modelo E-R presenta algunas limitaciones semánticas, y no es suficiente para expresar con detalle la estructura de algunos tipos de información. Por esta razón, surge el conocido como *modelo E-R extendido*, que amplía el modelo E-R añadiendo nuevos elementos. Con su mayor potencia, el modelo E-R extendido acerca el diseño conceptual a los conceptos de la programación orientada a objetos, incorporando por ejemplo mecanismos de herencia. El enfoque orientado a objetos recoge no solo la estructura del sistema de información, sino también su comportamiento dinámico.

Tras el diseño lógico, el diseño físico de la base de datos ha de llevar el modelo conceptual a la práctica y crear un repositorio de datos que pueda ser usado por el SGBD. Debe, asimismo, mantener todas aquellas propiedades del modelo conceptual, de modo que el contenido de la base de datos siga expresando de forma fiel la realidad y su estructura continúe siendo fácil de comprender para los usuarios. Si, siguiendo el enfoque más habitual, optamos por crear una base de datos según el modelo relacional, esto implica la

DNI	Nombre	Altura	Edad	Ciudad	Población	Superficie
50234561	Juan Gómez	1,85	35	Madrid	6386932	607
13254673	Edurne Montero	1,60	30	Toledo	80810	232
46576290	Luis Urrutia	1,75	46	Madrid	6386932	607
38941882	Juan Gomez	1, 71	55	Valencia	1564145	134

Cuadro 8.2: La información de las tablas `PERSONAS` y `CIUDADES` puede recogerse en una única tabla

creación de las correspondientes relaciones y los esquemas asociados a cada una de ellas.

La tablas que definamos en la base de datos pueden tener consecuencias directas sobre el uso de esta, afectando a aspectos como el rendimiento de las operaciones que posteriormente se lleven a cabo o al volumen de datos total necesario. Por ejemplo, nuestra base de datos con dos tablas, `PERSONAS` y `CIUDADES`, puede implementarse utilizando únicamente una tabla como la mostrada en el cuadro 8.2. Esta tabla contiene la misma información que las dos tablas anteriores, y en principio permite realizar operaciones similares. Si quisiéramos saber la población de la ciudad donde ha nacido una persona en concreto, podríamos hacerlo de igual modo con independencia de cuál de las estructuras mostradas tenga la base de datos. En un caso deberemos acudir a dos tablas y una interrelación entre ellas, mientras que en el otro solo es necesario emplear una tabla, la única que por otra parte contiene nuestra base de datos.

Aunque la funcionalidad sea la misma, el uso de una única tabla tiene efectos poco deseados que se advierten rápidamente, como por ejemplo la redundancia de datos. La población y superficie de Madrid aparecen repetidos en dos ocasiones, y aparecerían más veces si hubiera en la tabla `PERSONAS` más tuplas correspondientes a individuos nacidos en esta ciudad. De igual modo sucedería con otras ciudades. En el esquema basado en dos tablas, sin embargo, estos datos aparecen en una única ocasión y no dependen del número de personas de cada ciudad. En una base de datos de pequeñas dimensiones como la que utilizamos de ejemplo, esta circunstancia puede parecer poco relevante, pero si trabajamos con millones de registros en la tabla `PERSONAS` la diferencia es realmente importante.

Otro aspecto a tener en cuenta en el diseño físico de la tabla es elegir nombres adecuados para los atributos y las tablas. Los nombres deben ser inequívocos y dar una idea clara de la información que contienen, y un usuario debe poder identificar sin dificultades qué tablas y atributos son aquellos a los que debe acudir para efectuar una consulta y dónde se encuentra la información que busca. El atributo `CIUDAD` en la tabla `PERSONAS`, por ejemplo, cumple sin problemas su papel a la hora de establecer la relación entre esta tabla y la que recoge los datos de las distintas ciudades, pero si buscamos exclusivamente información sobre las personas, no es completamente preciso, ya que no aclara si se trata de la ciudad en la que una persona ha nacido o en la que habita.

Siempre que pueda existir alguna duda razonable a la hora de interpretar el contenido de una tabla, debe intentarse solventar esta mediante el uso de nombres claros y concisos. Establecer una sistemática a la hora de nombrar atributos y respetarla a lo largo de todo el conjunto de tablas de una base de datos hará más fácil para los usuarios la comprensión de esta. Por ejemplo, es habitual emplear el prefijo *num* cuando un atributo representa un conteo (y por tanto, su tipo de dato será

de tipo entero). Siguiendo esta convención, si quisiéramos añadir un campo a la tabla `PERSONAS` con el número de hermanos de cada individuo, sería más conveniente y más informativo denominar al atributo correspondiente `numHermanos`, en lugar de, por ejemplo, `Hermanos`. Más que seguir unas u otras normas para nombrar atributos y tablas, lo importante es ser consistente y tratar siempre de utilizar nombres que informen y no den lugar a confusiones.

Una vez que se establece un diseño y se implementa en la base de datos, lo normal es que este sea relativamente estable y no varíe a lo largo del tiempo. Las relaciones, por su parte, sí se modifican frecuentemente, ya sea añadiendo tuplas a medida que se incorporan nuevos datos o modificando las ya existentes. No obstante, los SGBD ofrecen también funcionalidades para modificar la estructura de la base de datos, incorporando nuevas tablas o cambiando el esquema de alguna de ellas. Estas funcionalidades no suelen ser accesibles para los usuarios con carácter general, sino pensadas para el mantenimiento de la base de datos por parte de su administrador.

8.2.7. Bases de datos espaciales

Todo cuanto hemos visto en los puntos anteriores constituye el conjunto de ideas fundamentales sobre las que se asienta la creación y uso de bases de datos de cualquier índole. No obstante, no hemos mencionado a lo largo de los ejemplos presentados ningún dato de carácter espacial, a pesar de que sabemos bien que la información geográfica contiene tanto una componente temática como una espacial. Más aún, algunos de los atributos en los sencillos casos mostrados, como puede ser el atributo `CIUDAD`, son fácilmente asociables a elementos geográficos (por ejemplo, un punto que señale el centro de la ciudad o un polígono que recoja su contorno).

Aunque las ideas anteriores no pierden su validez al incorporar datos espaciales, la inclusión de estos no es en absoluto obvia, y presenta una complejidad adicional que requiere de nuevos planteamientos para poder seguir trabajando con la base de datos de una forma similar a como sucede cuando se trabaja con los tipos de datos habituales. Mantener las características propias del SGBD en el contexto de los datos espaciales no es sencillo, y tampoco lo es integrar esa base de datos dentro de un SIG y permitir que este aproveche la potencia de dicha base de datos de la mejor manera posible.

En lugar de adentrarnos en la complejidad de las bases de datos espaciales (aunque en el capítulo 12 veremos bastante más en lo que a las operaciones y posibilidades de estas respecta), veremos las distintas etapas que podemos encontrar a lo largo de la historia de los SIG en lo referente a su integración con bases de datos, para de este modo comprender los diversas soluciones que han ido apareciendo.

8.3. Evolución del uso de bases de datos en los SIG

Como acabamos de decir, los conceptos que hemos visto en las anteriores secciones representan una gran parte de la realidad actual en cuanto al manejo de datos (espaciales o no) dentro de un SIG. No obstante, el problema del acceso a los datos se ha solucionado de diversas formas a lo largo de la historia de los SIG, y encontramos en las aplicaciones SIG distintos enfoques a lo largo del tiempo. Para concluir este capítulo veremos con algo más de detalle la evolución que ha seguido esta importante faceta de los SIG.

8.3.1. Primera generación. Ficheros

Los primeros programas, entre los cuales se han de incluir los primeros SIG, se caracterizaban en lo que al almacenamiento de datos respecta por una ausencia completa de cualquier tipo de almacenamiento estructurado. En estas aplicaciones, los datos no se veían como un elemento más dentro de un sistema, sino como una parte del propio *software* o, al menos, como algo asociado únicamente a un producto particular. Así, encontramos en esta época como práctica habitual el uso de ficheros con formatos cerrados, pensados para ser leídos y escritos casi de forma exclusiva por la aplicación particular que ha de consumirlos, limitando así el uso compartido y el alcance de los datos a otros ámbitos distintos.

Integrar en el SIG otros datos distintos a aquellos para los que la aplicación se había diseñado no era sencillo, ya que existía una vinculación muy directa entre software y datos. Asimismo, las funcionalidades del software eran también específicas para esos datos, y todas ellas se implementaban directamente en la aplicación. Al no existir un SGBD que se encargara de gestionar las operaciones, era el propio SIG quien debía ser responsable de las funcionalidades de acceso o edición. Otras funcionalidades típicas de un SGBD, sin embargo, no aparecían en estos primeros SIG, ya que no eran necesarias. Por ejemplo, velar por la integridad de los datos en operaciones concurrentes de varios usuarios no era necesario si la aplicación en sí no estaba diseñada para permitir este acceso múltiple.

Las únicas ventajas que pueden encontrarse en este enfoque son las relacionadas con el rendimiento, que podía en ciertos casos ser mayor que el esperable en caso de utilizar un SGBD para canalizar el trabajo con los datos. Esto es así debido a que la propia especificidad de la aplicación permitía una optimización «a medida», aunque todo ello a cambio de sacrificar la flexibilidad de la aplicación, su escalabilidad, o la posibilidad de que los datos empleados pudieran ser utilizados de forma sencilla para alimentar otras aplicaciones.

8.3.2. Segunda generación. Bases de datos relacionales

Una vez que las bases de datos comienzan a tomar su papel en el panorama del software, no tardan en encontrar su camino dentro de las aplicaciones SIG. Las bases de datos relacionales, que como ya sabemos son las más empleadas, comienzan a ser utilizadas también para gestionar los datos espaciales con los que se trabaja en un SIG. A partir de esta segunda generación, se empiezan a adaptar las características del modelo relacional y de las bases de datos que lo implementan a las particularidades de los datos espaciales. Las dificultades que aparecen debido a la inherente complejidad de la componente espacial hacen que surjan diversas alternativas para su manejo. Las más reseñables de entre ellas son el uso de una arquitectura dual en la que únicamente la componente temática se gestiona mediante una base de datos y el uso de una arquitectura en capas en el que se da un pleno almacenamiento de la información espacial en la base de datos.

Arquitectura dual

El primer intento de incorporar las bases de datos lo encontramos en el uso de una arquitectura dual en la cual el SGBD se hace cargo únicamente de la componente temática de los datos. Puesto que la dificultad estriba en el manejo de la componente espacial, esta no se incorpora por el momento a la base de datos, que trabajará únicamente con los datos temáticos. Esto permite el uso de sistemas gestores de bases de datos estándar, sin adaptación alguna, ya que estos se encuentran perfectamente preparados para el ma-

nejo de esos datos no espaciales, y no requieren elementos adicionales para trabajar sobre ellos.

La componente espacial, por su parte, es gestionada por el propio SIG, en el que se implementan las funcionalidades necesarias. Al igual que sucedía anteriormente con los SIG de primera generación, no todas las funcionalidades de un SGBD han de aparecer necesariamente, ya que el sistema encargado de permitir el trabajo con los datos no es como tal un SGBD. La única diferencia reside en que en este caso esta circunstancia afecta tan solo a la componente espacial de los datos, mientras que la componente temática queda en manos de un verdadero SGBD.

Existen, por tanto, dos subsistemas encargados de la gestión de los datos, cada uno de los cuales se encarga de un tipo de información (Figura 8.6). Esta arquitectura en la que datos espaciales y datos no espaciales se encuentran separados tiene ciertas ventajas, puesto que permite reutilizar información ya existente de uno u otro tipo. Por ejemplo, ficheros procedentes de aplicaciones CAD pueden incorporarse en el SIG aunque carezcan de una componente temática, aprovechando, no obstante la información espacial. El resultado de este uso es en su mayoría de tipo gráfico, pero un SIG que presente una arquitectura dual puede trabajar con él y gestionarlo gracias al subsistema encargado de la información espacial, suponiendo ya una mejora respecto al enfoque de los SIG de primera generación.

La división entre datos espaciales y no espaciales conlleva, no obstante, una serie de inconvenientes. Por un lado, resulta difícil integrar operaciones en las que se empleen ambas componentes de los datos, que requerirán llamadas a ambos subsistemas y la posterior combinación de sus respuestas. Toda esta labor debe implementarse en el SIG, siendo este un proceso costoso que complica el desarrollo. Si todo el manejo de datos recayera sobre la base de datos, estas operaciones se realizarían de forma transparente, ya que bastaría ejecutar la operación en el SGBD y este se encargaría de realizar las tareas pertinentes y devolver después al SIG la respuesta. Se evitaría asimismo la redundancia en el propio software, ya que al emplear dos subsistemas han de duplicarse una buena parte de funcionalidades, una de ellas en el SGBD externo y otra en el propio SIG.

Aunque una parte importante del SIG descansa ya sobre un SGBD, otra sigue presentando muchas de las deficiencias que caracterizaban a la primera generación. Mientras que la componente temática disfruta de las ventajas de usar un SGBD, la componente espacial no goza aún de esas ventajas, y existe una cierta descompensación que limita las posibilidades y hace más complejo el desarrollo del sistema.

Arquitectura en capas

La otra forma de aprovechar una base de datos relacional para su uso dentro de un SIG consiste en incorporar toda la información dentro de la base de datos, incluyendo la de corte espacial, buscando la manera más adecuada de llevar esto a cabo pese a las limitaciones que la propia base de datos presenta en este caso. Asumiendo que una base de datos relacional en su concepto tradicional no esta diseñada para contener objetos complejos tales como geometrías o imágenes, y que, especialmente, el SGBD correspondiente no presenta las mismas funcionalidades y la misma potencia en el manejo de este tipo de datos que en el de tipos de dato estándar (valores numéricos, cadenas de texto, fechas, etc.), es posible, sin embargo, plantear soluciones que permitan llevar toda la información de un SIG a una base de datos y poder gestionarla por completo a través de un SGBD, con las ventajas que ello conlleva, y que ya conocemos.

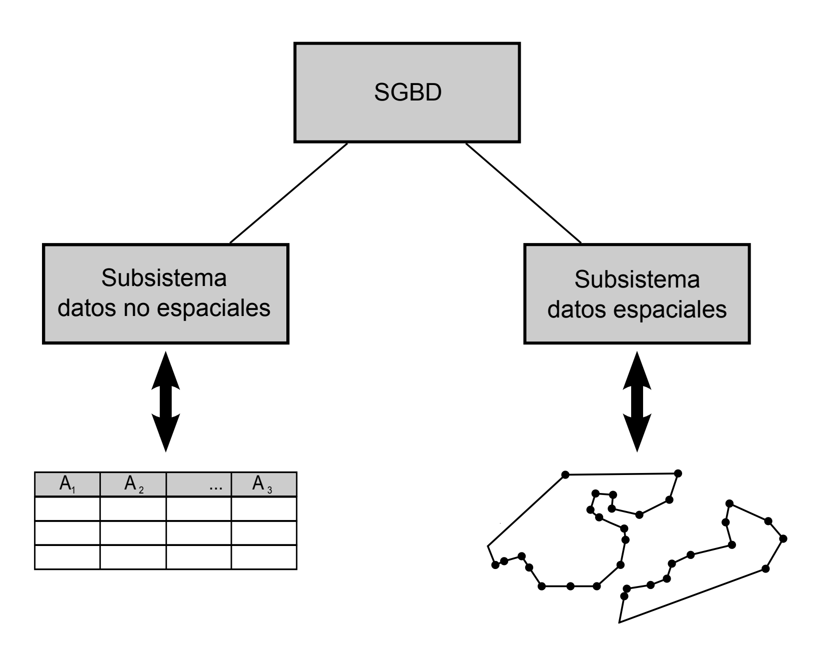


Figura 8.6: Arquitectura dual con subsistemas distintos para el manejo de datos espaciales y no espaciales.

Dos son las alternativas existentes: un almacenamiento *transparente* y un almacenamiento *opaco*. Ambos se distinguen en la forma de almacenar la información y también las operaciones sobre los datos, que vienen condicionadas por la estrategia empleada para el almacenamiento de estos.

En el almacenamiento transparente se emplean los propios tipos de datos del SGBD, y las operaciones se implementan en el lenguaje de consulta de este. Es decir, se intenta implementar toda la funcionalidad deseada empleando los elementos básicos del SGBD de la misma forma que haríamos si los datos a almacenar no fueran de tipo espacial. La componente espacial de los datos se almacena empleando tuplas, variando según la implementación la manera en que esto se lleva a cabo. Una geometría como tal no se ajusta a ningún tipo básico de datos, pero en realidad esa geometría no es sino un conjunto de coordenadas que definen una serie de puntos, y dichas coordenadas sí que son un tipo básico susceptible de almacenarse en un SGBD común.

En el almacenamiento opaco se emplean objetos binarios para almacenar la información y las operaciones se implementan externamente en la herramienta SIG. Al no utilizar los tipos de datos del SGBD, tampoco pueden emplearse las operaciones de consulta de este, y es necesario implementar los algoritmos correspondientes en el SIG.

La ventaja más directa de utilizar una arquitectura en capas, ya sea mediante un almacenamiento transparente o uno opaco, es la facilidad para reutilizar un SGBD existente. Con poco esfuerzo pueden incorporarse los datos espaciales a un SGBD estándar, existiendo en la actualidad numerosas alternativas sobradamente probadas y con una amplia gama de funcionalidades. Esta es la opción más empleada hoy en día en los SIG, principalmente por esa sencillez, que permite una conexión sin muchas dificultades de una

aplicación SIG con la mayoría de los SGBD de uso habitual fuera del ámbito SIG.

Existen, aun así, inconvenientes y aspectos mejorables, achacables a la nula especialización de los SGBD para el manejo de información espacial. En el caso del almacenamiento opaco, no poder emplear el lenguaje de consulta del SGBD constituye un grave inconveniente. Por su parte, en el almacenamiento transparente sí que puede emplearse, pero no todas las operaciones necesarias para el trabajo con datos espaciales pueden implementarse con un lenguaje de consulta no adaptado a las particularidades de los datos espaciales, por lo que la funcionalidad es limitada.

Asimismo, la eficacia es menor, ya que en un caso los algoritmos son externos al SGBD y en el otro las consultas suelen ser complejas y operan sobre un elevado número de tuplas, necesario para recoger la información espacial.

8.3.3. Tercera generación. Bases de datos extensibles

En la actualidad, las bases de datos presentan arquitecturas extensibles que permiten adaptarse a la naturaleza de los datos con los que trabajan, de tal forma que enfocan sus funcionalidades hacia la tipología particular que se maneje. Los tipos de datos clásicos conviven con nuevos tipos de datos que pueden ser definidos, y con operaciones específicas para estos.

Un caso particular de estas bases de datos extensibles son las *bases de datos orientadas a objetos*, que ya fueron comentadas al presentar los distintos modelos de bases de datos. A pesar de que este tipo de bases de datos no ocupan una porción significativa en el mercado global de las bases de datos, y son las de tipo relacional las más extendidas, existen algunos sectores en los que han logrado una mayor penetración, entre ellos el del SIG. Por sus características, las bases de datos orientadas a objetos resultan ventajosas para el manejo de datos complejos que no puedan recogerse con facilidad utilizando los tipos de datos clásicos de una base de datos relacional. En este grupo pueden incluirse las primitivas geométricas que utilizamos en un SIG para recoger la componente espacial de un dato espacial, las cuales resulta más adecuado considerar como objetos de un tipo dado (punto, línea o polígono), aprovechando así las ventajas que un enfoque orientado a objetos proporciona.

La principal ventaja de una base de datos orientada a objetos es su mayor eficiencia en el acceso a datos, lo que se traduce en consultas más rápidas en comparación con una base de datos relacional (veremos más sobre consultas en bases de datos espaciales en el capítulo 12). Por el contrario, carece de la base matemática de esta, por lo que el soporte para esas consultas es menos robusto. Para saber más sobre bases de datos orientadas a objetos, puede consultarse [79].

Los SGBD actuales presentan en su gran mayoría extensiones dedicadas al manejo de datos espaciales, los cuales contienen todo lo necesario para el manejo óptimo de estos, la realización de ciertas operaciones fundamentales y la optimización de las consultas y operaciones. Esta optimización es posible ya que el tipo de datos espacial está plenamente integrado en la base de datos y es considerado de la misma manera que cualquiera de los tipos de datos estándar como puede ser una cadena de texto o un valor numérico. La eficiencia que se obtiene de este modo es muy elevada.

8.4. Resumen

En este capítulo hemos visto los conceptos básicos sobre bases de datos. Una base de datos constituye un sistema que permite un manejo adecuado de los datos, garantizando la seguridad

e integridad de estos y permitiendo el acceso a distintos usuarios de forma transparente. La base de datos está formada por los datos en sí, organizados de forma estructurada, mientras que las operaciones las provee el sistema gestor de base de datos (SGBD).

Existen diversos modelos para el almacenamiento de datos, siendo el modelo relacional el más habitual en la actualidad. En el modelo relacional la información se organiza en tablas relacionadas entre sí. Cada fila de una base de datos conforma una tupla, que contiene la información correspondiente a una entidad dada.

El diseño de la base de datos es de gran importancia, y conlleva el diseño de un modelo conceptual, el diseño de un modelo físico, la implementación y el mantenimiento. Herramientas como los diagramas E-R son de ayuda en las fases de diseño, cuyo principal objetivo es crear una estructura de la base de datos que facilite la interpretación de la información contenida y permita el máximo aprovechamiento de esta.

En lo que a los SIG respecta, las bases de datos se han ido incorporando paulatinamente a la gestión de los datos espaciales. Partiendo de una situación inicial en la que no se empleaban sistemas gestores de bases de datos, estos han ido integrándose en los SIG de diversas formas. En la actualidad, se emplean bases de datos relacionales adaptadas para poder almacenar datos espaciales y realizar operaciones sobre ellos. Los SGBD extensibles representan la ultima tendencia, y en ellos puede integrarse plenamente la información geográfica de forma óptima.

Capítulo 9

Metadatos

Blake, Landon; Olaya, Víctor

9.1. Introducción

Los datos contienen la información geográfica, y es esta la que empleamos en un SIG para realizar las distintas operaciones que ya hemos visto en capítulos anteriores de este libro. No obstante, esos datos pueden resultar insuficientes, ya que el proceso de interpretación mediante el que extraemos la información a partir de estos puede requerir conocer otra serie de elementos.

Por ejemplo, si tenemos las coordenadas de un punto disponemos de un dato, pero para interpretarlo correctamente necesitamos conocer, entre otras cosas, el sistema de coordenadas en que vienen expresadas esas coordenadas. El dato con el que trabajamos (las coordenadas), requiere unos datos adicionales (por ejemplo, el código EPSG del sistema de referencia empleado) para cobrar verdadero sentido.

Surge así el concepto de *metadatos*. Literalmente, los metadatos son «datos acerca de los datos», y su misión es *explicar* el significado de los datos. Es decir, ayudan a los usuarios de los datos a entender mejor el significado que estos tienen y la información que guardan. Los metadatos son un documento adicional que acompaña a los datos, y que permite una mejor gestión y una utilización más precisa de ellos.

Trabajando en el entorno de un SIG, los datos con los que trabajamos son de tipo espacial, y como ya estudiamos en su momento (véase el apartado 4.3), existe una componente espacial y una temática. Los metadatos pueden referirse a ambas componentes, ya que es necesario documentar todas ellas, y podemos encontrar metadatos referidos a una capa de forma global, a su componente espacial o a su componente temática.

Un ejemplo de metadato global de una capa puede ser el nombre de su autor o la fecha en la que ese dato ha sido creado. El sistema de referencia en el que se expresan las coordenadas de cada entidad recogida es un tipo de metadato relativo a la componente espacial. Y en lo referente a la componente temática, los metadatos pueden recoger las unidades en las que se recoge una variable asociada a cada entidad, o bien almacenar cualquier otro valor que permita una mejor interpretación de esa variable.

En una definición más formal, los metadatos son archivos de información que recogen las características básicas de algún dato o recurso. Representan el quién, qué, cuándo, dónde, cómo y por qué de ese recurso. Los metadatos geoespaciales se emplean para documentar recursos geográficos digitales tales como una base de datos espacial, un SIG o una imagen de satélite. Un

registro de metadatos incluye elementos básicos tales como el título o nombre del recurso, elementos geográficos como la extensión que cubre el dato o el sistema de coordenadas empleado, así como elementos relativos a la base de datos asociada tales como la definición de cada uno de sus campos o el dominio en que se encuentran los valores de estos.

El concepto de metadato no es algo nuevo y exclusivo de los datos digitales, ya que un mapa impreso también contiene metadatos en cierta forma. Una leyenda o un texto en un margen del mapa con información sobre la fecha en que se ha creado son también metadatos. En el caso de los datos geográficos digitales, los metadatos no forman parte del dato directamente, sino que son independientes de este. Ello permitirá realizar operaciones separadamente con los metadatos, tales como búsquedas, que abren nuevas posibilidades y dan un gran valor a estos.

9.2. La utilidad de los metadatos

Dependiendo del tipo de dato con el que trabajemos y las operaciones que deseemos realizar con ellos, los metadatos correspondientes serán más o menos necesarios, pudiendo ser prácticamente irrelevantes o bien completamente imprescindibles. Por ejemplo, si se trabaja con una única capa y gran parte de la información que esta contiene no va a emplearse para la realización de operaciones, los metadatos son menos necesarios que si se da un uso más intenso a los datos.

En algunos casos, incluso si carecemos de metadatos, resulta posible interpretar correctamente los datos, como sucede si trabajamos con un Modelo Digital de Elevaciones (MDE) y valores de elevación en metros. Es fácil saber que los valores de elevación se encuentran en esas unidades aplicando cierta lógica, y procesarlos correspondientemente aunque no exista un dato explicito que así nos los indique.

En otras circunstancias, los metadatos son necesarios, pues contienen información que no puede inferirse directamente desde los propios datos. Si varias capas están en sistemas de coordenadas distintos y deseamos aplicar las transformaciones correspondientes para unificarlos en uno único y procesarlas de manera conjunta, estas transformaciones no se pueden llevar a cabo si no conocemos el sistema de origen del que partimos en cada capa. En este supuesto, el trabajo con los datos viene condicionado a que existan los metadatos correspondientes.

Los metadatos son, por tanto, sumamente importantes en el trabajo con SIG y, como veremos en breve, cobran una importancia mayor todavía cuando no nos encontramos en el contexto de un uso aislado de los datos, sino cuando nos situamos en un entorno de un gran volumen de datos y numerosos usuarios.

Dos de las funciones principales de los metadatos son garantizar el uso correcto y adecuado de los datos y facilitar su gestión, localización y consulta.

9.2.1. Garantizar el uso correcto de los datos

Uno de los beneficios más importantes que proporcionan los metadatos es asegurar que los datos espaciales son empleados de forma adecuada. Los datos espaciales, como muchos otros datos, son creados habitualmente para un determinado objetivo, y este objetivo no ha de ser necesariamente evidente o contenerse como tal en los datos mismos. Cuando se emplean esos datos para un objetivo distinto a aquel para el que fueron diseñados, pueden surgir problemas debido a que se está realizando un proceso para el que los datos con los que se trabaja presentan

carencias. A continuación veremos algunos ejemplos para ilustrar algunas situaciones en las que puede producirse un uso indebido de los datos, las cuales podrían corregirse o evitarse mediante el empleo de los correspondientes metadatos.

1. Un organismo crea un juego de datos con los ejes de las principales vías de una ciudad. Estos datos se emplean para labores de mantenimiento, de tal modo que faciliten la localización de las señales viales en la realización de inventarios. El juego de datos no contiene informacion sobre direcciones ni tampoco almacena la topología de la red. Posteriormente, una compañía especializada en reparto adquiere estos datos para el cálculo de rutas óptimas desde sus almacenes hasta las direcciones de destino de sus clientes.
2. Un organismo crea un juego de datos con los elementos de la red de alcantarillado tales como alcantarillas, tuberías, bombas, etc. Durante años, esta capa no se actualiza. Años después de la creación de esos datos, ese mismo organismo desarrolla un proyecto relativo a la calidad de las aguas y el control de la contaminación y utiliza ese juego de datos.
3. Una compañía mantiene un registro de los límites aproximados de las parcelas catastrales de una zona. El juego de datos, no obstante, no muestra las posibles discrepancias que pueden existir en esos límites, tales como solapes o huecos. Una inmobiliaria emplea ese juego de datos para asesorar a sus clientes y mostrarles la localización y límites de las parcelas a los compradores potenciales.

En los tres casos, nos encontramos con un usuario de un juego de datos que, por desconocer las características de este, realiza un uso indebido.

En el primer caso, la compañía de reparto no podrá operar con esos datos, ya que no contienen la información que necesitan. Conocer de antemano las limitaciones de los datos antes de adquirirlos o plantear cualquier operación con ellos ahorra tiempo y dinero.

En el segundo caso, la información contenida en el juego de datos está desfasada. Si los metadatos contienen la fecha en la que los datos fueron creados, esta puede emplearse para juzgar la validez de estos últimos en función de su antigüedad.

Por último, en el tercer caso la compañía inmobiliaria trabaja con unos datos que no tienen la precisión requerida. Si existieran metadatos y estos dejaran claro que los límites de parcelas son aproximados, se conocería con exactitud las limitaciones de los datos y no se les daría un uso indebido.

Estos tres ejemplos ponen de manifiesto la necesidad que existe de conocer acerca de los datos más que lo que ellos mismos contienen, en particular todo lo relativo a los fines para los que estos se han creado. Esto permite conocer lo que se puede esperar de los datos y no emplearlos en situaciones indebidas.

Los creadores de datos deben procurar acompañar estos de metadatos precisos y suficientes, y los usuarios deben consultarlos antes de utilizar dichos datos. De este modo, se puede garantizar que un dato no se utiliza de forma errónea y que los resultados que se obtendrán tendrán validez.

Vimos en el capítulo 7 cómo la calidad se define como el conjunto de propiedades y de características de un producto o servicio que le confieren su aptitud para satisfacer unas necesidades explícitas e implícitas. Los metadatos documentan esas características y las de todas aquellas necesidades a las que pueden responder los datos, y de este modo documentan la propia calidad del dato. Como ya se dijo entonces, los metadatos son un elemento muy importante en relación con la calidad de los datos espaciales

9.2.2. Facilitar la gestión los datos

Las funciones anteriores ponen de manifiesto la utilidad e importancia de los metadatos en un contexto reducido en el que un individuo o un pequeño grupo trabaja con ciertos datos. La importancia de los metadatos se hace patente incluso cuando se dispone de un único dato (una sola capa), pues es en el momento de utilizar este cuando se consultan los metadatos y se emplea la información que contienen para poder conocer la precisión de los datos, su referencia espacial, u otros elementos que permitan que ese uso sea más correcto.

En esta situación, se dispone ya de los datos y de los metadatos, y estos últimos nos permiten conocer más acerca de los primeros. No obstante, en el panorama tecnológico actual un usuario no dispone de todos los datos que necesita, sino que puede acceder a ellos en la medida en que le sea necesario, del mismo modo que no guardamos en nuestro ordenador enciclopedias y libros, pero podemos acceder a ellos a través de Internet. Las tecnologías que veremos en el capítulo 24 dedicado a servidores y clientes nos permiten acceder a una enorme cantidad de datos espaciales, y los metadatos juegan un papel clave en la gestión de estos.

Los metadatos cobran una importancia mayor si cabe a la hora de organizar los datos espaciales dentro de ese contexto tecnológico, ya que informan de las características de los datos a sus multiples usuarios. Los metadatos constituyen un «resumen» de las características principales de los datos, y pueden ser empleados para labores de búsqueda y localización de datos de un tipo dado. De este modo, compartir los datos es más sencillo, y la difusión de estos se realiza de una forma más fluida.

Algunas características de los datos solo se contienen en los metadatos, como por ejemplo el sistema de coordenadas empleado o la descripción detallada de los distintos campos de la componente temática. Otros, por el contrario, pueden obtenerse a partir de los propios datos, como por ejemplo el área que cubre una capa. Aunque este tipo de valores sea posible obtenerlos procesando los datos en sí, añadirlos a los metadatos abre nuevas posibilidades en el marco de la gestión, permitiendo un manejo más dinámico.

En general, los metadatos son mucho menos voluminosos que los datos a los que acompañan. Si buscamos datos para una zona dada, es mucho más sencillo consultar los metadatos que consultar los datos como tales. Mientras que la primera operación puede realizarse de forma rápida, la segunda requiere unos cálculos mucho mayores. Con grandes volúmenes de datos, esa búsqueda puede ser virtualmente irrealizable. Es decir, que los metadatos facilitan y agilizan la localización de los datos cuando estos se buscan por criterios geográficos. Añadiendo a los metadatos elementos como la extensión del área cubierta por los datos, este tipo de búsquedas se efectúan de forma más ágil y efectiva.

Cuando la búsqueda se realiza por otros criterios distintos, los metadatos son el elemento clave para poder realizar esta búsqueda. Si queremos localizar la capa más actual con un tipo de información dada, necesitamos conocer *qué* información contiene cada capa y *cuándo* se ha creado, para aplicar sobre esos datos los criterios de búsqueda correspondientes. Sin los metadatos, estas operaciones no son posibles.

En su conjunto, los metadatos sirven para catalogar los datos y por tanto son básicos para hacer más fluida la transferencia de estos.

Facilitando la localización de datos adecuados para una determinada tarea se obtienen además beneficios colaterales. Haciendo más sencillo el acceso a los datos se pueden evitar esfuerzos redundantes tales como la creación o modificación de datos cuando ya existen otros que pueden ser-

vir para responder a una necesidad concreta. El uso de metadatos permite así ahorro de tiempo y dinero y un mejor aprovechamiento de los datos.

9.3. Características de los metadatos

Los metadatos pueden ser tan variados en sus características como los propios datos a los que acompañan. Los enfoques para la creación de metadatos son muy diversos y ello da lugar a metadatos muy diferentes.

Algunas de las características que resulta de interés tratar son las siguientes:

- **Contenido** de los metadatos. ¿Qué información contienen?
- **Granularidad** de los metadatos. ¿A qué elementos particulares hace referencia esa información?
- **Forma de almacenamiento** de los metadatos. ¿Cómo se guardan?

9.3.1. Contenido de los metadatos

Los valores que pueden incorporarse a los metadatos son muy abundantes, tantos como tipos distintos de información se considere necesario registrar respecto a un dato geográfico particular.

Las características de los metadatos asociados a los datos dependerán directamente de estos y de algunos factores como los siguientes:

- **El tipo de dato** y, en particular, **el modelo de representación empleado**. Los datos vectoriales tendrán asociados unos metadatos distintos que los correspondientes a datos ráster.
- **El formato en que se almacenan los datos**. El tipo de fichero o base de datos condiciona la información que puede almacenarse (vimos esto en detalle en la sección 6.9), y por tanto condiciona los metadatos.
- **La organización, entidad o individuo responsable** de la creación de los datos y el uso que se pretende dar a estos. Puesto que, como hemos dicho, los datos se crean para un objetivo definido, este objetivo y los intereses de quien ha creado los datos definirán el tipo y cantidad de información que se recoja en los metadatos. Datos pensados para un catálogo público tendrán asociados metadatos distintos que datos privados con acceso restringido, del mismo modo que datos pensados para un uso muy concreto presentarán unos metadatos diferentes a los que acompañarán a unos datos de uso más genérico.
- **El elemento al que se asocian los metadatos**. Como veremos en el siguiente apartado, podemos asociar metadatos a un juego de capas, una capa o una entidad aislada dentro de una capa. Esto implica diferencias en el contenido de los metadatos, pues esos elementos tienen características de distinta naturaleza.
- **El estándar empleado** para crear los metadatos. En el apartado 24.7 veremos los estándares que existen para los metadatos geográficos y la forma que estos tienen, la cual define directamente su contenido.

Algunos de los elementos comunes que se incorporan a los metadatos geográficos son los siguientes:

- **Información de identificación**. Permite identificar de forma única un dato geográfico y distinguirlo de otros. Esta información ayuda a catalogar los datos, e incluye el nombre, palabras clave, una descripción básica o la ya mencionada extensión geográfica de los datos.
- **Información sobre la calidad de los datos**. Puede incluir, entre otros elementos, aquellos relativos a la completitud de estos, los procesos que se han empleado en su creación y mantenimiento, o las operaciones de validación y verificación a las que se han sometido.

En relación con los procesos empleados, es importante reseñar que muchos de los datos geográficos provienene de la aplicación de algoritmos sobre datos de partida. que se utilizan para crear datos geográficos algún tipo de dato geográfico. Para documentar la calidad de esos datos resultantes, se debe documentar en los metadatos la procedencia completa de estos, indicando las metodologías empleadas para su creación y todos los metadatos propios de las capas de entrada.

Un ejemplo de esto puede ser el proceso de cálculo de una capa de pendientes a partir de un MDE. Este MDE tendrá a su vez unos metadatos asociados (entre ellos algunos relativos a su calidad), y la bondad y calidad de la capa de pendientes está ligada directamente a la del MDE. Por tanto, en los metadatos debe hacerse referencia a ese MDE o bien a las características de este.

Si el MDE no se ha adquirido directamente, sino que se ha elaborado haciendo uso de otros procesos tales como interpolación a partir de curvas de nivel, se ha de añadir también a los metadatos la información correspondiente a esos procesos, especificando por ejemplo el método de interpolación usado, los parámetros de ajuste de este o incluso el software mediante el que se ha aplicado.

Con esto, puede «rastrearse» el origen de los datos y se dispone de una base sobre la que evaluar la calidad de estos en función de dicho origen. Tenemos así el concepto de *linaje* de los datos. Esta idea es similar a la de *trazabilidad* empleada en otros sectores como, por ejemplo, el alimentario.

- **Información sobre la representación del dato espacial**. Se incluyen en este grupo la precisión y exactitud de los datos, la escala de trabajo o la resolución en el caso de capas ráster. Este tipo de metadatos están también íntimamente ligados con la calidad de los datos.
- **Información sobre la componente no espacial**. Información relacionada con los atributos que acompañan a las capas vectoriales, o bien relativas a las variables que se recogen en capas ráster. Esto incluye explicaciones sobre el significado de los nombres de los atributos, el rango de valores válidos para cada uno de ellos o los métodos empleados para recoger estos datos.
- **Información sobre la distribución**. Esta información sirve para definir el acceso a los datos y las posibilidades de distribución de estos, especificando quiénes pueden acceder a ellos y quiénes no, o en qué condiciones pueden hacerlo. También puede recoger elementos como la fecha en que fueron publicados los datos o bien cuándo fueron puestos a disposición del público, de tal forma que se disponga de toda la información referente a su presencia en el marco de una Infraestructura de Datos Espaciales.

De entre estos, algunos son considerados como fundamentales y se incluyen de forma genérica, mientras que otros pueden o no incorporarse. Al definir una especificación de metadatos, se pueden establecer niveles de prioridad, estableciéndose un grupo de propiedades básicas que han de documentarse siempre y otro con propiedades de carácter opcional.

9.3.2. Granularidad de los metadatos

Habitualmente, los metadatos están asociados a un conjunto de datos al completo. Este conjunto de datos que sirve como unidad a la hora de crear metadatos coincide en general con la idea de capa en un SIG. Es decir, cada capa tiene asociado un bloque de metadatos.

Esto no quiere decir, no obstante, que no puedan registrarse metadatos a un nivel distinto. Dependiendo del tipo de datos con los que se trabaje, puede resultar de interés o incluso necesario asociar metadatos a unidades diferentes.

Algunos metadatos como el sistema de coordenadas serán compartidos por todos los elemen-

tos de una capa, y por tanto es lógico en su caso emplear la capa como unidad básica en lo que a metadatos se refiere. Otro metadatos, sin embargo, hacen referencia a elementos particulares dentro de la capa.

Este tipo de metadatos aparecen especialmente cuando, a lo largo del ciclo de vida de los datos, se introducen modificaciones en estos, editándolos o añadiendo nuevas entidades. Si bien en el origen el creador de los datos es una única entidad, otras entidades pueden alterar esos datos y deberán actualizar correspondientemente los metadatos. Registrando como autores de los datos a ambas entidades se recoge más información al respecto, pero esta puede no ser suficiente. Sabemos que los datos fueron creados por una entidad A y posteriormente modificados por una entidad B, pero si tomamos un elemento dado no podemos saber si esta corresponde al trabajo original de A o a la modificación realizada por B.

De modo similar, podemos incorporar a los metadatos las dos fechas de creación y edición de los datos, así como parámetros relativos a la calidad de los datos o las metodologías empleadas para recogerlos en ambos instantes. Sin embargo, no podemos saber en qué fecha fue incorporado un elemento concreto o la calidad de los datos que definen ese elemento en particular.

En estas circunstancias, resulta más conveniente optar por metadatos más granulares, de forma que puedan recogerse particularizados para las distintas entidades de la capa.

Por ejemplo, algunos de los datos que pueden resultar de interés a escala de elemento (en el caso de una capa vectorial, hablamos de una geometría y sus atributos asociados) son los siguientes:

- Quién ha creado ese elemento.
- Quién ha modificado ese elemento.
- Cuándo fue creado originalmente.
- Cuándo ha sido modificado por ultima vez.
- Cuántas veces ha sido modificado.
- Una descripción del objeto real que este elemento representa.

Podemos encontrar el caso opuesto, en el que varias capas comparten los mismos metadatos, y por tanto estos pueden asociarse a escala de toda una familia de datos. Ese es el caso cuando se tiene un conjunto de capas generadas por una misma entidad y para un mismo fin, las cuales cubren una amplia zona geográfica y debido a ello se encuentran divididas horizontalmente. Estas circunstancias se dan de forma habitual en series de datos de carácter nacional o autonómico, y conforman una de las situaciones en las que el registro de metadatos puede hacerse para toda la serie en su conjunto, al menos para algunos de esos metadatos.

Los metadatos pueden así registrarse a una escala distinta a la de la capa como unidad de datos, aunque esta sigue siendo la referencia más habitual a la hora de crear metadatos (Figura 9.1).

9.3.3. Forma de almacenamiento de los metadatos

Si para los propios datos geográficos encontramos muy diversas alternativas a la hora de almacenarlos, la situación no es distinta a la hora de almacenar los metadatos. Las dos alternativas principales son el uso de ficheros independientes o el almacenamiento en bases de datos [80].

Elegir entre uno u otro enfoque depende del conjunto de datos de trabajo, su volumen total, el uso principal que se le da o la granularidad de los datos según vimos en el apartado anterior. Se recomienda el uso de bases de datos cuando los datos estén sujetos a frecuentes modificaciones o si existe una parte de los metadatos que es común a varios grupos de datos. Este es el caso que vimos en el apartado anterior al mencionar los

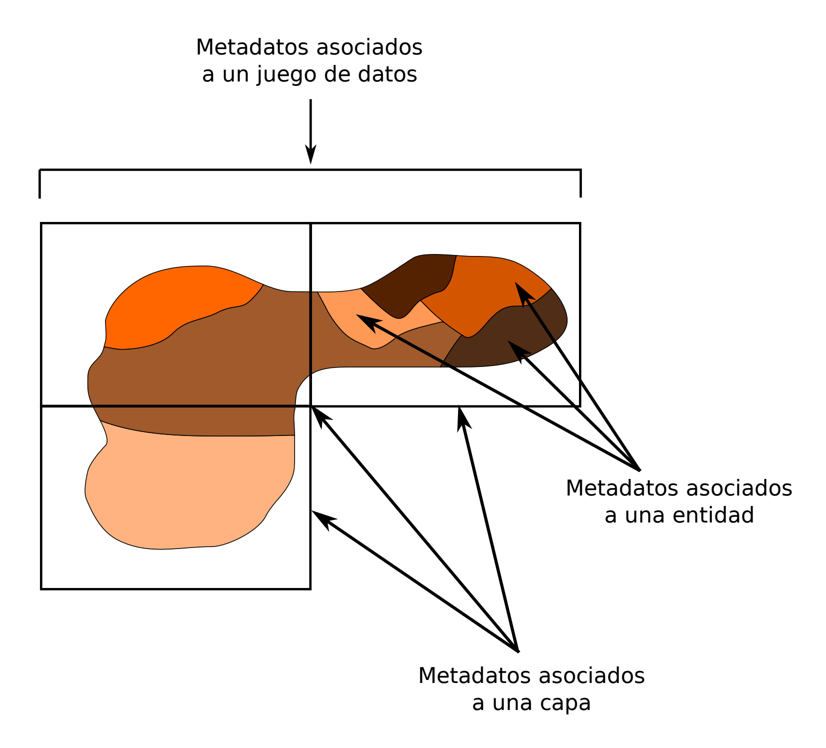


Figura 9.1: Granularidad de los metadatos. Los metadatos pueden hacer referencia a elementos a distinta escala.

metadatos correspondiente a toda una serie de datos.

Utilizando una base de datos, resulta más sencillo actualizar estos, especialmente si puede haber varios usuarios que realicen esas actualizaciones. Las modificaciones que estos usuarios hagan han de reflejarse en los metadatos, y para ello es necesario contar con una tecnología que permita un acceso concurrente similar para los metadatos. Las bases de datos proveen esas capacidades, y son por tanto adecuadas para el almacenamiento de metadatos en ese contexto.

Si, por el contrario, los datos no van a ser usados de esa forma, no es probable que deban modificarse con frecuencia y apenas contienen elementos comunes, una forma más simple de almacenarlos es utilizando ficheros independientes, generalmente ficheros de texto plano que son más sencillos de producir y además pueden leerse con un simple editor de texto.

9.4. Creación de metadatos

La creación de los metadatos no es tarea de un único grupo de profesionales ni se lleva a cabo en un único momento dentro del ciclo de vida de los datos. Por el contrario, distintas entidades o grupos pueden crear o editar los metadatos, y pueden hacerlo a lo largo de todo el tiempo de existencia de dichos datos.

Los metadatos puede crearse en el mismo origen de los datos, recogiendo la información al mismo tiempo que se producen los datos en sí. Esta creación puede derivar de la digitalización de mapas impresos o de la medición directa de valores, entre otros procesos. Las organizaciones

que se encargan de crear datos son responsables en este caso de crear los metadatos que los acompañan.

Las entidades responsable de distribuir datos geográficos y ponerlos a disposición de los distintos usuarios pueden igualmente crear metadatos en caso de que estos no existan. Estas entidades no producen datos, pero recogen datos de sus creadores y han de prepararlos para ofrecer un mejor servicio. Los metadatos aportan un valor añadido a los datos y facilitan la gestión de datos que estas organizaciones han de realizar.

Por último, los mismos usuarios y beneficiarios de los datos pueden crear metadatos o ampliar los ya existentes. Si estos usuarios modifican los datos, esas modificaciones deben recogerse en los metadatos. Aún así, incluso si no se producen modificaciones, puede resultar de interés añadir nueva información en los metadatos, particularmente aquella que estos no contengan pero que pueda tener valor para los objetivos que se persiguen con su uso. Igualmente, dejar de usar los datos por alguna razón tal como el hecho de que se encuentren desactualizados es una información que los mismos usuarios pueden incorporar a los metadatos, informando así a futuros usuarios de la falta de validez de esos datos.

En resumen, los metadatos pueden ser creados o modificados en los siguientes puntos dentro de la vida de los datos:

- Cuando se crean los datos.
- Cuando se organizan o catalogan los datos.
- Cuando se modifican o editan los datos.
- Cuando se archivan o descatalogan los datos.

En circunstancias ideales, todo dato debería tener asociados unos metadatos, y estos últimos deberían crearse siempre que se creen los primeros y actualizarse siempre que estos se modifiquen. La realidad, sin embargo, es que una gran parte de los datos geográficos que existen no tiene metadatos asociados, o bien estos no son lo suficientemente detallados.

Una razón importante para ello es la falta de concienciación que existen por parte tanto de creadores de datos como de usuarios respecto a la importancia de los metadatos. Mientras que para un usuario aislado o un pequeño grupo de técnicos SIG puede no resultar importante generar metadatos a la hora de crear algún dato geográfico, cuando nos encontramos con organizaciones más grandes e infraestructuras de datos mayores, los metadatos se hacen imprescindibles. El usuario aislado prefiere generalmente no dedicar tiempo (la creación de metadatos no es en absoluto sencilla y es una tarea que consume tiempo) a crear unos metadatos que no percibe como importantes para su trabajo con los datos.

Si en lugar de datos geográficos habláramos de libros, una persona normal no cataloga los libros que tiene en su casa y recopila información acerca de cada uno de ellos, almacenándola en una base de datos. En una gran biblioteca, sin embargo, esta labor es imprescindible, pues de otro modo resulta imposible gestionar tanto el gran fondo bibliográfico del que se dispone como el amplio número de lectores y usuarios.

En realidad, incluso en el nivel más local, las ventajas de la creación de metadatos son grandes, especialmente si consideramos que un dato creado y utilizado en un entorno local puede más adelante pasar a formar parte de una infraestructura de datos de mayor envergadura.

La creación de metadatos no tiene que ser necesariamente una labor propia del técnico o equipo de técnicos que crean los datos en sí, del mismo modo que el escritor de un libro no es el encargado de catalogar este. Tanto usuarios como creadores de datos geográficos deben poseer unos conocimientos básicos en relación a los metadatos, pero existen expertos en metadatos a

quien la creación de estos debe corresponder en última instancia.

Los usuarios deben saber consultar e interpretar los metadatos, y ser conscientes de la importancia de estos y el papel que juegan en una buena parte de las operaciones que pueden desarrollarse con los datos. Los creadores, por su parte, deben ser capaces de elaborar no los metadatos en sí directamente, pero sí la información necesaria acerca de los datos que debe incluirse en los metadatos, y transmitirla de forma correcta a los profesionales encargados de crear estos.

9.4.1. Herramientas para crear metadatos

Existe un amplio conjunto de herramientas que facilitan la labor de creación de metadatos. Entre ellas podemos distinguir las siguientes [80].

- **Editores de texto**. Los metadatos pueden almacenarse en un fichero de texto plano, y por tanto pueden editarse con cualquier programa que permita la creación y edición de tales ficheros. Lo habitual en este caso es disponer de un fichero plantilla que contenga los distintos campos que se han de registrar para cada conjunto de datos geográficos, y la creación del metadato consiste simplemente en apoyarse en esa plantilla y añadir el valor correspondiente a cada campo.
- **Formularios**. A partir de una definición de campos como la anterior, se pueden desarrollar herramientas más elaboradas que presenten una interfaz gráfica con distintas cajas de texto o listas desplegables. Estas aplicaciones, además de ser más agradables para el usuario, permiten incorporar elementos de validación en el proceso, evitando que en algún campo se introduzcan valores incorrectos o avisando al usuario en caso de que un campo presente un valor sospechoso. Del mismo modo, se puede establecer qué campos son obligatorios y cuáles opcionales, y avisar en caso de que un metadato no contenga valores para todos sus campos obligatorios.
- **Utilidades**. Existen aplicaciones que no se emplean directamente para introducir los valores de los metadatos, pero que pueden intervenir en el proceso. Entre ellas están aquellas que chequean y validan los metadatos o las que lo preprocesan dándole un formato adecuado según unas reglas establecidas de antemano.
- **Herramientas de creación automática de metadatos**. Algunos de los valores que se incorporan a los metadatos pueden extraerse de los propios datos. Por ello, el proceso de creación de metadatos puede automatizarse en cierta medida, y existen aplicaciones específicamente diseñada para realizar esa tarea.
  Las aplicaciones de creación automática de metadatos pueden, por ejemplo, analizar un archivo con una capa vectorial y crear un archivo adjunto de metadatos en el que se incluya la extensión de la capa, el tipo de geometrías que tiene o los campos de su tabla de atributos, indicando además el tipo de valor en cada uno de ellos.
  Además de estos metadatos extraídos directamente del dato geográfico, las herramientas que automatizan este proceso pueden añadir información común introducida manualmente en una única ocasión, y que se repite de forma automática en todos los datos creados. Así, por ejemplo, si una de estas herramientas automáticas se emplea en un organismo, puede establecer como creador de cada nuevo dato a esa entidad, sin necesidad de que la persona encargada de crear dicho dato deba añadir esa información manualmente cada vez que genere algo nuevo.
  El uso de herramientas automáticas no se limita al momento de creación de los datos, sino que también pueden emplearse durante la actualización de estos. Si se actualiza un dato empleando

un SIG, este puede estar conectado con la aplicación de creación automática de metadatos y lanzar esta para que vuelva a analizar ese dato y actualizar los metadatos inmediatamente. Quizás sea necesario añadir información manualmente, pero una buena parte de esta habrá sido creada de forma automática, facilitando el proceso y haciéndolo más rápido.

La importancia de este tipo de aplicaciones es grande si se tiene en cuenta que, como se ha dicho, una de las razones principales de la carencia de metadatos es la cantidad de tiempo que se requiere para elaborarlos.

9.5. Resumen

Para que los datos sean verdaderamente útiles es necesario acompañarlos de otros datos adicionales que los describan y aporten información suplementaria acerca de ellos. Estos datos adicionales son los metadatos, y recogen información tanto de la componente espacial como de la componente temática del dato geográfico.

La importancia de los metadatos se hace patente en la gestión de la calidad de los datos, o a la hora de utilizarlos como base para procesos, pues es necesario conocer todos los detalles relativos a los datos con los que se trabaja. Asimismo, permiten que las operaciones de descubrimiento y consulta de los datos se efectúen de forma eficaz.

Los metadatos pueden asociarse con los datos geográficos en niveles de detalle diversos, desde una única entidad hasta una colección de varias capas. El contenido de los metadatos es también variable y depende de su granularidad, así como de otros parámetros, como por ejemplo el tipo de modelo datos.

Algunas de las secciones más importantes que encontramos en los metadatos son la identificación del dato, los valores relativos a su calidad o los relacionados con su distribución, entre otros.

Parte III

El analisis

Capítulo 10

¿Qué puedo hacer con un SIG?

10.1. Introducción

El análisis de los datos geográficos es, junto con la generación cartográfica, una de las tareas fundamentales sin las cuales el concepto de SIG no alcanza su verdadero significado. La información que manejamos en un SIG es a su vez una fuente de nueva información, y solo es a través de su análisis que podemos obtener esta última y sacar partido de ella.

Los datos espaciales contienen mucha más información de la que a primera vista nos muestran. Todo dato espacial es el resultado de un proceso localizado espacialmente, el cual podemos conocer en mayor medida si sabemos «leer» la información subyacente que dicho dato contiene. Los cursos de los ríos informan sobre la estructura del terreno o la litología existente. Los patrones en los que se disponen los nidos de una especie de ave dicen mucho acerca del comportamiento de esta. Y así, muchos otros ejemplos que pueden analizarse según alguna o varias de las formulaciones que veremos a lo largo de esta parte del libro. Lo importante es conocer qué formas hay de convertir la información espacial en información sobre un proceso dado, o cómo extraer parámetros de utilidad a partir de datos espaciales relacionados con nuestro área de estudio.

Ahora que ya conocemos cómo son los datos que podemos incorporar a un SIG, es el momento de ver cómo los SIG constituyen herramientas de gran potencia para estudiar dichos datos. El análisis de estos datos geográficos ha cobrado una nueva dimensión desde la aparición de los SIG, surgiendo nuevos planteamientos y mejorándose los ya existentes. A lo largo de toda su historia, el análisis ha sido uno de los elementos más importantes de un SIG, y a día de hoy existen formulaciones que cubren casi todo el abanico posible de necesidades. Dados unos datos espaciales y un problema a resolver, es probable que exista algún procedimiento que, si no nos da una solución como tal a partir de dichos datos, pueda emplearlos para acercarnos a esta. El objetivo de esta parte del libro es mostrar los más importantes de dichos procedimientos, con el fin de dejar clara la potencialidad de los datos espaciales en los más diversos contextos de estudio.

10.2. ¿Qué es el análisis espacial?

De modo formal, podemos decir que el análisis espacial es el estudio cuantitativo de aquellos fenómenos que se manifiestan en el espacio [81]. Ello indica una importancia clave de la posición, la superficie, la distancia y la interacción a través del propio espacio. Para que estos conceptos cobren sentido, se necesita que toda la información esté referenciada espacialmente.

En otros tipos de datos, el orden en que los elementos, indexados de una u otra forma, se dispongan no es relevante, y una reestructuración de los mismos no afecta necesariamente a los resultados. En el caso del dato espacial, este supuesto no se cumple, y alterar la distribución espacial de los valores de partida implica que los resultados de cualquier análisis espacial serán distintos. Esto nos permite definir tal análisis espacial como el conjunto de técnicas y modelos que hacen un uso explícito de la referencia espacial de cada dato particular [82].

Desde un punto de vista menos formal, podemos entender el análisis espacial sencillamente como el conjunto de operaciones que desarrollamos en base a los datos espaciales en el trabajo habitual con estos. En realidad, prácticamente todo cuanto hacemos con los datos espaciales constituye alguna forma de análisis. Pese a que, como ya sabemos, la aparición de los SIG ha propiciado un fuerte desarrollo en este terreno, el análisis espacial no es algo novedoso, y todos en algún momento hemos utilizado algún mapa clásico y efectuado algún análisis sobre él.

Por ejemplo, hemos cogido un mapa de elevaciones y hemos buscado dónde se sitúa el pico más alto, o bien hemos visto la elevación concreta a la que se encuentra un elemento dado como, por ejemplo, una población. Igualmente, hemos mirado en un mapa de carreteras intentando planificar una jornada turística, viendo qué lugares de interés podemos visitar o cómo llegar desde uno a otro de estos lugares haciéndolo por las mejores carreteras o de la forma más rápida. Estas actividades habituales son ejemplos de análisis geográficos que podemos igualmente realizar dentro de un SIG.

Cuando los datos espaciales que se recogen en ese mapa de elevaciones o ese mapa de carreteras pasan al marco de un SIG, se incorporan igualmente a un marco de análisis en el que además se contemplan otra serie de operaciones. Dentro de ese marco, se da por igual una formalización de los métodos de análisis espacial, fundamentada sobre todo en la naturaleza numérica y cuantitativa con la que se maneja todo dato espacial dentro de un Sistema de Información Geográfica. Esta formalización es la que provee el entorno adecuado para plantear todo tipo de análisis cuantitativos sobre los datos espaciales.

No debe verse el análisis espacial como un conjunto de complejos algoritmos, sino como una colección de procesos con los que explotar los datos espaciales. Estas operaciones producen resultados de diversa clase, como pueden ser los siguientes:

- Capas de datos geográficos, en cualquiera de los formatos posibles, y con cualquier tipo de información.
- Tablas de datos.
- Escalares o vectores.

En ocasiones, los resultados expresan la misma variable que el dato de partida (por ejemplo, el cálculo de una media), y en otros las variables de entrada y salida son distintas (por ejemplo, si a partir de una capa de elevaciones calculamos una de pendientes).

Asimismo, todo análisis espacial parte de un conjunto de datos espaciales, pudiendo estos ser de un único tipo, o de varios distintos que se combinan en un procedimiento concreto. Por ejemplo, en el caso de calcular la localización del punto más alto el resultado es una sencilla coordenada, y tan solo se utiliza la variable elevación. En el caso de la altura media de una ciudad, se utilizan dos entradas. Por un lado, la elevación, y por otro el emplazamiento de la ciudad. Aunque en ese mapa de carreteras que antes citábamos toda esa información se contiene en una única hoja, en realidad son dos elementos distintos combinados a la hora de representarlos. En términos más acor-

des con un SIG, podemos decir que tenemos dos capas distintas que utilizamos como entradas.

10.3. Razonamiento y consulta geográfica en un SIG

Tanto si la información geográfica se encuentra dentro de un SIG como si no, las distintas herramientas de análisis espacial permiten estudiar los datos geográficos con el fin de obtener respuestas a cuestiones muy variadas. Estas cuestiones son las que motivan el propio análisis espacial y el desarrollo de las formulaciones correspondientes.

A lo largo de esta parte del libro veremos una amplia serie de dichas formulaciones. Utilizando estas, obtenemos de los datos espaciales los valores que sirven como respuesta a las cuestiones planteadas. Sin embargo, la mayor importancia a la hora de plantear un análisis espacial apropiado no recae sobre los algoritmos, sino sobre los planteamientos previos. Es decir, no sobre las respuestas, sino sobre las preguntas que nos planteamos acerca de los datos y los procesos espaciales. Por ello, es importante desarrollar un adecuado razonamiento espacial y conocer bien el tipo de cuestiones que podemos plantear y cómo formular estas.

Según [83], pueden distinguirse las siguientes categorías de cuestiones geográficas:

- Relativas a posición y extensión
- Relativas a la forma y distribución
- Relativas a la asociación espacial
- Relativas a la interacción espacial
- Relativas a la variación espacial

Por su parte, [84] cita la siguiente lista como compendió básico de las cuestiones que pueden plantearse en el análisis espacial.

- ¿Dónde se encuentra?
- ¿Dónde tiene lugar?
- ¿Qué hay ahí?
- ¿Por qué esta ahí?
- ¿Por qué no está en otro lugar?
- ¿Qué podría haber ahí?
- ¿Podría estar en otro lugar?
- ¿Cuánto hay en ese lugar?
- ¿Hasta dónde se extiende?
- ¿Por qué tiene esa estructura particular?
- ¿Existe regularidad en esa estructura?
- ¿De qué naturaleza es esa regularidad?
- ¿Por qué el patrón de distribución espacial tiene que mostrar regularidad?
- ¿Dónde se encuentra en relación a otros de su misma clase?
- ¿Qué clase de distribución conforma?
- ¿Se encuentra en todo el mundo?
- ¿Es universal?
- ¿Donde están sus límites?
- ¿Cual es la naturaleza de esos límites?
- ¿Por qué esos límites acotan su distribución?
- ¿Qué otras cosas encontramos asociadas espacialmente a ese fenómeno?
- ¿Aparecen esas cosas asociadas espacialmente en otros lugares?
- ¿Por qué deben estar asociadas espacialmente?
- ¿Ha estado siempre allí?
- ¿Cuándo apareció por primera vez?
- ¿Cómo ha variado espacialmente a través del tiempo?
- ¿Qué factores han influido en su dispersión?
- ¿Por qué se ha dispersado de esa manera?
- ¿Qué factores geográficos han limitado su dispersión?

Con todo lo anterior, un SIG es una herramienta de primer orden tanto para formular cuestiones geográficas como para resolverlas.

Por ejemplo, a través de la mera representación de los datos geográficos, un SIG puede ayudar a identificar y definir los problemas a plantear, ya que la exploración visual de los datos es un elemento clave en la formulación de interrogantes

geográficos. Esa misma representación puede ser también empleada para dar respuesta a dichos interrogantes, puesto que esas respuestas quedan muchas veces patentes con el simple análisis visual. Considerando la gran variedad de formas en que los datos espaciales pueden representarse dentro de un SIG, ello hace que la visualización sea de por sí una componente de los SIG de gran interés para el análisis.

Asimismo, mediante las capacidades de superposición de capas y la visualización de dicha superposición se pueden tratar en un SIG las cuestiones relativas a la asociación entre distintas variables espaciales, y las que conciernen a la variación temporal de estas variables pueden estudiarse con conjuntos de representaciones sucesivas a modo de «instantáneas». También en esto las funcionalidades de visualización de la información espacial son herramientas de indudable utilidad.

A lo anterior deben sumarse los distintos algoritmos de análisis geográfico, que iremos viendo a lo largo de los próximos capítulos, y que aprovechan la arquitectura particular de un SIG para explotar en profundidad la información espacial.

10.4. Tipos de análisis espacial

A la hora de analizar los datos en busca de respuestas a cuestiones como las anteriores, existen muchos enfoques distintos.

Aunque dar una taxonomía del análisis espacial es difícil y la casuística es excesivamente amplia, podemos establecer algunas divisiones principales. Para cada una de ellas, veremos algunos ejemplos, con el objetivo de cubrir distintos supuestos y, una vez más, mostrar en detalle qué tipo de operaciones pueden realizarse a partir de datos espaciales. Estas divisiones no pretenden ser un conjunto exhaustivo ni una clasificación formal, sino simplemente hacer ver la variedad de análisis posibles y su complejidad.

Debe resaltarse que la implementación de estos distintos tipos de análisis, así como la de las formulaciones más detalladas que iremos viendo en sucesivos capítulos, varía de unos SIG a otros. Nuevamente, aquí se tratan como elementos teóricos y su utilización práctica en un SIG dado puede ser de una forma u otra. Lo importante es conocer ese elemento conceptual y saber *qué* podemos obtener a partir de un determinado dato espacial, para poder así estudiar un problema concreto y plantear una forma de resolución del mismo.

Las siguientes son algunas de las familias principales en las que englobar los procedimientos de análisis:

- **Consulta espacial**. El tipo de análisis más simple es la consulta directa de la información que contienen los datos espaciales. Es el tipo de análisis básico que usamos cuando trabajamos con cartografía clásica, y nos proporciona información inmediata a partir de una simple observación de los datos. La propiedad fundamental que empleamos es la posición de cada elemento geográfico.
  Este tipo de operaciones da respuestas a preguntas de tipo
  ¿Qué tipo de suelo encontramos en una coordenada (x, y) dada?
  ¿Dónde se encuentra la localidad x?
  Son, como puede verse, análisis que se pueden resolver simplemente «mirando» al mapa que contiene la información de partida, y por ello constituyen la forma más sencilla de análisis espacial.
  Puesto que en la información geográfica dentro de un SIG todo elemento tiene asociadas unas propiedades en forma de valores, también podemos consultar estos valores. Así, podemos plan-

tear consultas no necesariamente relacionadas con la componente espacial, tales como

¿Cuáles son las diez ciudades españolas con mayor población?

¿Qué pueblos de España comienzan por la letra A?

Combinar este tipo de consultas con las puramente espaciales constituye un análisis sencillo pero fundamental dentro de los posibles en un SIG, y representa una de las utilidades más frecuentemente empleadas de estos en el trabajo diario.

- **Análisis topológico**. Las consultas hechas a las capas de datos espaciales pueden tener relación no solo con su posición sino con la relación con otros elementos de la misma capa. La existencia de topología (ver 5.3.2) puede emplearse para la realización de consultas que respondan a cuestiones como, entre otras, las siguientes:

  ¿Cómo llegar desde mi posición actual hasta una coordenada concreta por la red viaria existente?

  ¿Qué comunidades autónomas comparten límite con Madrid?

- **Medición**. La existencia de una referencia espacial para cada uno de los elementos con los que trabajamos en el análisis dentro de un SIG hace que podamos cuantificar otra serie de parámetros también espaciales. El más básico de estos parámetros es la distancia, que puede ser una distancia simple entre dos puntos dados o bien una distancia entre elementos complejos tales como polígonos o líneas, o combinaciones de ellos.

  Además de la distancia podemos medir otras propiedades tales como

  Área
  Perímetro
  Longitud de un recorrido no lineal
  Factores de forma

  Dentro de este grupo incluimos parámetros más elaborados tales como pendientes, o índices diversos que derivan todos ellos de medidas sencillas similares a las anteriores. Estas medidas no tiene que ser necesariamente de tipo espacial, ya que conceptos como la pendiente pueden medirse no solo sobre un espacio geográfico —variación de z sobre el plano xy— sino también sobre otras variables —variación de dicha variable (temperatura, concentración de un nutriente, etc.) sobre el plano xy—.

  Responden a preguntas muy variadas tales como

  ¿Qué superficie de zonas arboladas hay en mi término municipal?

  ¿Cuántos kilómetros comprende la red viaria española?

  ¿Tienen las distintas zonas de usos de suelo formas compactas o por el contrario son principalmente alargadas y de tipo fusiforme?

- **Combinación**. Uno de los procedimientos más habituales y más característicos dentro del uso de un SIG es la combinación o superposición de varias capas de información. La propia estructura de la información geográfica en capas facilita notablemente estos procedimientos.

  Antes de la existencia de los SIG, la combinación de capas implicaba la utilización de mapas en soportes tales como transparencias o acetatos, una opción farragosa y muy poco apta para el análisis de las combinaciones resultantes. Dentro de un SIG, existen metodologías para integrar la información de varias capas en formas muy distintas, y las nuevas capas resultantes pueden luego analizarse con sencillez independientemente de su origen.

  La estructura de las bases de datos geográficas es idónea para integrar toda la información disponible acerca de una región geográfica concreta, y las distintas capas que forman esta se pueden combinar de forma sencilla tanto para su análisis como para su simple visualización.

- **Transformaciones**. Podemos englobar dentro de este grupo una amplia serie de procedimientos que modifican los elementos de entrada de diversas formas.

  Por ejemplo, uno de los procedimientos más frecuentes dentro de un SIG es la creación de áreas de influencia. Este tipo de operaciones de análisis convierte los distintos elementos geográficos en áreas que reflejan la influencia de dicho elemento en base a parámetros tales como distancias o costes. Se tiene así una transformación geométrica, ya que la forma del objeto se transforma en una nueva que indica la zona que se ve afectada por dicho objeto.

  Con ellas podemos responder a preguntas como

  ¿Qué puntos de la ciudad no tienen una farmacia a menos de un kilómetro de distancia?

  ¿Están los distintos comercios de un barrio demasiado juntos, de forma que probablemente estén compitiendo por la clientela?

  Si considero que para una escapada de fin de semana el turista medio recorre como mucho 100 kilómetros, ¿qué municipios alrededor del mío son susceptibles de venir de visita turística y por tanto debería promover en ellos los valores naturales de este?

  Otros ejemplos de este tipo de modificaciones geométricas es la simplificación de líneas, que trata de definir los mismos trazados de un conjunto de líneas reduciendo el número de puntos empleados.

  También se pueden realizar transformaciones de las geometrías en función no solo de su componente espacial (sus coordenadas), sino utilizando igualmente los valores asociados a estas. Un ejemplo de esto es la agrupación de geometrías que comparten algún atributo común en entidades únicas. Dado un conjunto de polígonos con los distintos términos municipales, para los cuales exista un atributo que indique la comarca a la que pertenecen, se pueden agrupar estos para obtener polígonos únicos de cada comarca.

  Otras transformaciones son de tipo cartográfico, tales como la conversión entre sistemas de coordenadas distintos, las reproyecciones, o la aplicación de transformaciones afines en general. Estas son básicas para, por ejemplo, combinar datos referenciados según distintos sistemas.

  Un tipo de transformación importante es la relativa a los modelos de datos, pues estos, como ya sabemos, son tan variados como los sistemas de coordenadas. Las transformaciones entre formatos de almacenamiento son importantes para un manejo óptimo de los datos geográficos, ya que ciertas operaciones se realizan de manera más adecuada en unos formatos concretos. Igualmente, la combinación de capas requiere en muchos casos que estas se encuentre en un mismo modelo de datos, al igual que sucede con los sistemas de coordenadas. La conversión entre los modelos ráster y vectorial, la interpolación o el cálculo de capas de densidad son ejemplos de análisis que modifican la forma de representación de una realidad espacial concreta.

  Por último, encontramos transformaciones basadas en los valores de las variables estudiadas. Dentro de este grupo encontramos las reclasificaciones, que en el caso de datos categóricos transforman la identificación de cada elemento en una clase dada, o los cambios de escala u otras operaciones aritméticas tales como la normalización de una variable en un rango dado, o la tipificación de una variable para asimilar su distribución de valores a la de una curva normal. Estos últimos se efectúan sobre datos de tipo continuo.

- **Análisis de superficies**. El análisis de superficies es uno de los más potentes de cuantos encontramos en un SIG. Desde parámetros básicos como la pendiente o la orientación hasta parámetros morfométricos muy específicos, pasando

por todas las herramientas del análisis hidrológico, la batería de operaciones disponibles es muy amplia. Aunque este análisis de superficies se entiende como el de la superficie terrestre (es decir, el relieve), gran parte de estas operaciones pueden aplicarse a cualquier otra superficie, entendiendo esta en su sentido matemático. Así, la pendiente indica una tasa de variación y puede aplicarse a capas con valores distintos de la elevación, tales como temperaturas, densidades, etc.

- **Estadística descriptiva**. Los elementos de la estadística clásica tienen sus equivalentes en los datos espaciales, y nos permiten calificar cuantitativamente los datos con los que trabajamos. Se incluyen aquí descriptores de centralidad y dispersión, de dependencia espacial o el estudio de patrones espaciales, entre otros muchos. Estos pueden a su vez usarse para el contraste de hipótesis que contengan una cierta componente espacial.
Por ejemplo, estos estadísticos nos permiten dar respuesta a cuestiones del tipo

    ¿Es constante la media de altura a lo largo de toda la geografía de mi país?

    ¿Existe alguna tendencia de los individuos de una especie a congregarse, o por el contrario se dispersan por todo el territorio disponible minimizando el contacto con otros congéneres?

    ¿Existe alguna dirección predominante en los movimientos de individuos de una especie o se desplazan erráticamente?

- **Inferencia**. Otro análisis estadístico de gran importancia en los SIG es el que permite inferir comportamientos de las distintas variables y estudiar, por ejemplo, la forma en que estas van a evolucionar a lo largo del tiempo.
El establecimiento de modelos de cambio y variación representa una de las herramientas más actuales en el campo de los SIG, y un campo en abundante desarrollo.

- **Toma de decisiones y optimización**. La realización de actividades en el medio tiene una obvia componente espacial. Son muchos los parámetros que influyen en ellas, y en función de estos dichas actividades se desarrollarán de una forma u otra. La estructura de la información geográfica en capas dentro de un SIG, favorable como ya vimos para la superposición de capas, lo es igualmente para estudiar de forma combinada los efectos de distintos factores.
El estudio de estos factores puede ser una herramienta clave para tomar decisiones relativas a la actividad sobre la que ejercen su influencia. Así, los procedimientos de análisis espacial nos sirven para responder a cuestiones como, por ejemplo,

    ¿Cuál es el mejor lugar para emplazar una nueva construcción en función de su impacto sobre el medio?

    ¿Por qué trazado es más conveniente construir una nueva carretera?

    ¿Dónde situar un nuevo hospital para que el servicio en la comarca mejore lo máximo posible?

Dentro de estos análisis, muchos de ellos tratan de maximizar o minimizar alguna función objetivo dependiente de los factores implicados, que pueden ser tanto variables recogidas en distintas capas como parámetros espaciales tales como distancias.

- **Modelización**. La creación de modelos espaciales dentro de un SIG es una tarea aún pendiente de mucho desarrollo. No obstante, existe un gran número de modelos en los más diversos campos, y la arquitectura de datos y procesos de los SIG es propicia para la implementación de otros nuevos.
Modelos como los de tipo hidrológico son habituales en los SIG más populares, y la estructura raster de los datos que se emplean generalmente en estos facilita en gran medida el análisis y la

implementación de modelos distribuidos. Otros modelos que encuentran en los SIG una plataforma idónea para su implementación son los basados en autómatas celulares, con aplicación en muchas áreas distintas.

Como ya se ha dicho, todos estos tipos de análisis no son independientes entre sí, y la verdadera potencia de un SIG radica en la elaboración de metodologías que los combinen. Por ejemplo, la elaboración de áreas de influencia considerando distancia a través de una red viaria, utilizando la topología de esta, que incorpora el estudio de la topología de la red, la medición de distancias sobre la misma, y la transformación de entidades geográficas en función de lo anterior.

10.5. Resumen

En líneas generales, todo cuanto hacemos con la información geográfica implica algún tipo de análisis. Desde una mera consulta a un modelo muy complejo, este análisis explora dicha información y permite obtener resultados que descubren otros tipos de información subyacente.

Existe una gran variedad de procesos de análisis espacial. Estos pueden tomar datos espaciales de diversas clases y generar resultados también muy diversos, por lo que su clasificación es compleja. Algunos de los más característicos de cuantos podemos llevar a cabo dentro de un SIG son aquellos que sacan partido de la forma en que este maneja las distintas capas de información. Por ejemplo, la superposición de capas o el análisis combinado de distintos factores como herramienta de apoyo en la toma de decisiones.

En este contexto, deben considerarse los SIG como herramientas que van a permitir una mejor formulación de las cuestiones geográficas, y que del mismo modo van a ayudar en la búsqueda de respuestas a estas.

Conceptos básicos para el análisis espacial

11.1. Introducción

Trabajar con datos espaciales tiene una serie de implicaciones que han de considerarse con detenimiento antes de llevar a cabo cualquier análisis. A lo largo de esta parte del libro veremos formas muy distintas de analizar los datos espaciales para obtener resultados de índole variada, y todas ellas tienen en común el hecho de trabajar sobre este tipo particular de datos. Conocer en profundidad el dato espacial es, por tanto, imprescindible, no solo en lo relativo a su forma, su manejo y su almacenamiento —que ya fue visto en la parte correspondiente— sino también en lo referente a su análisis y cómo ha de tratarse e interpretarse la información que contiene —que lo veremos en el presente capítulo.

El objetivo de este capítulo es múltiple. Por una parte, presentar las principales particularidades de los datos espaciales, así como la manera de tener estas en cuenta a la hora del análisis. Por otra, estimular un correcto razonamiento espacial y un entendimiento adecuado tanto de las limitaciones como de la potencialidad de los datos espaciales como fuente del análisis geográfico. Y por último, presentar algunos de los fundamentos teóricos sobre los cuales se crean después todas las metodologías de análisis, las estadísticas espaciales y los algoritmos que se detallarán en los capítulos sucesivos.

Estos fundamentos incluyen algunas nociones básicas sobre matemática del plano y el espacio, y conceptos sobre las posibles relaciones existentes entre objetos geográficos.

11.2. Particularidades de los datos espaciales

Considerar que el dato espacial es un dato cualquiera sin ninguna peculiaridad supone no realizar sobre él un análisis óptimo. Las características propias de los datos espaciales dotan a estos de una gran potencialidad de análisis, al tiempo que condicionan o limitan otras operaciones. Asimismo, estas particularidades son el origen de una gran parte de los retos aún existentes dentro del análisis geográfico, y por sus implicaciones directas no pueden desestimarse sin más. Su conocimiento es, por tanto, imprescindible para todo tipo de análisis espacial.

El carácter especial del dato espacial deriva de la existencia de posición. Esta posición se ha de entender tanto en términos absolutos (posición de una entidad en el espacio expresada por sus coordenadas) como relativos (relación con otras entidades también en dicho espacio). Las consecuencias de que todo dato espacial se halle por definición localizado a través de coordenadas son diversas, y deben enfocarse desde los distintos puntos de vista del análisis espacial.

Algunos de los puntos que a continuación veremos representan problemas que han de tenerse presentes en el análisis. Otros son simplemente conceptos básicos que deben conocerse pero no han de implicar necesariamente una dificultad asociada.

11.2.1. Escala

En el apartado 3.4 vimos con detalle el concepto de escala cartográfica, y cómo este se aplica de igual modo a la representación y gestión dentro de un SIG. Existe, además, otra forma de considerar la escala, y que resulta de especial interés para los contenidos de esta parte: la escala de análisis.

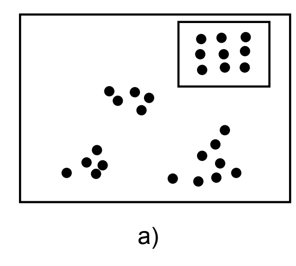
A la hora de estudiar la información geográfica, podemos hacerlo a distintos niveles y, dependiendo del nivel elegido, los resultados serán de una u otra naturaleza. Esto se manifiesta en las estructuras espaciales (véase más adelante en esta misma sección), que condicionan los valores que se derivan de sus análisis a través de las distintas formulaciones de análisis. Este hecho es fácil verlo con algunos ejemplos, que nos permitirán comprobar cómo a distintas escalas los datos geográficos tienen características distintas.

Por ejemplo, sea el conjunto de puntos de la figura 11.1. En el ejemplo a) se ve que los puntos se agrupan en conglomerados en zonas concretas del espacio. Esto es lo que denominaremos una estructura *agregada*. Sin embargo, si nos acercamos y solo enfocamos uno de dichos grupos, el de la parte superior derecha —ejemplo b)—, la estructura que vemos claramente no responde a una estructura agregada, sino que los puntos se disponen más o menos equiespaciados. Es lo que se conoce como estructura *regular*. Dependiendo de a qué escala observemos y analicemos la estructura espacial del conjunto de puntos, esta resulta de un tipo o de otro.

La escala de análisis debe ir inseparablemente relacionada con el fenómeno que pretendemos analizar, ya que es esta la que le da sentido. Supongamos el caso de llevar a cabo un análisis del relieve. Dependiendo de a qué escala observemos dicho relieve, la imagen que obtenemos es muy distinta. A un nivel global, distinguimos las grandes cadenas montañosas, y el resto del relieve aparece más o menos llano. Si nos acercamos a alguna de esas zonas llanas, se aprecia un relieve que antes no percibíamos, con ondulaciones y accidentes orográficos de menor entidad, que son suficientes para apreciarse a esta escala, pero no a la escala global anterior. Siguiendo este proceso, podemos ir acercándonos progresivamente hasta que incluso un pequeño grano de arena constituya un relieve notable.

Si vamos a llevar a cabo un estudio de cómo el relieve influye en los movimientos de las masas de aire a nivel de todo el planeta, no tiene sentido estudiar las formas del relieve a este último nivel de máximo detalle. Como se muestra en la figura 11.2, si para definir las formas de relieve en un punto dado lo hacemos considerando dicho punto y los valores de elevación a su alrededor, la caracterización que hagamos varía en función de la dimensión de esa zona alrededor (que es la que define la escala de análisis). Para valores pequeños de dicha zona de análisis, el punto analizado puede definirse como una cima, mientras que aumentando la escala de análisis se advierte que el punto se sitúa en el fondo de un valle.

Por tanto, debemos observar el relieve desde la distancia correcta a la cual la información que nos proporciona es la más adecuada para un análisis dado. Además de existir una escala de mayor relevancia para un análisis concreto, lo cierto es que el conjunto de todas las escalas de análisis contiene en su totalidad una información más amplia que la correspondiente a una única



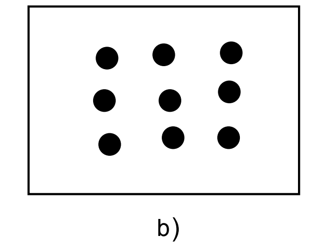


Figura 11.1: Dependiendo de la escala de análisis, la estructura de un conjunto de puntos puede ser distinta.

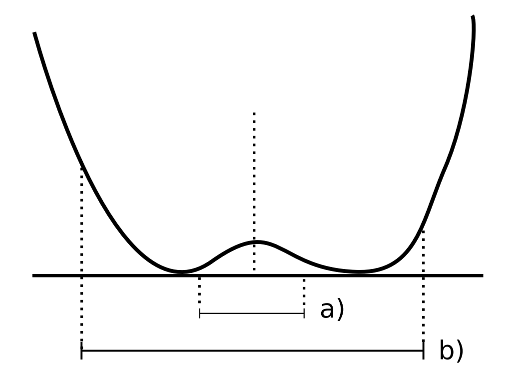


Figura 11.2: Dependiendo de la escala de análisis, un mismo relieve puede ser caracterizado como cima (a) o fondo de valle (b)

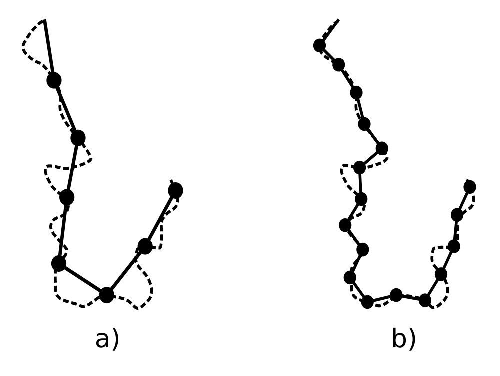


Figura 11.3: La unidad de medida empleada modifica el resultado obtenido.

escala, y por tanto resulta de interés el trabajar a múltiples escalas y combinar los resultados.

Este enfoque de escalas múltiples es relevante también en relación con los propios datos, independientemente de lo que representan. Es decir, independientemente de la escala y la dimensión «real», y en relación solo con la escala definida por el formato de los mismos. Por ejemplo, en el caso de imágenes, el uso de operadores a diferentes escalas (referida aquí la escala al número de píxeles utilizados en el operador) es ventajoso para realizar ciertas operaciones tales como la detección de bordes [85](véase 17.3.2). Combinado esto con lo anterior, la importancia de la escala en el análisis espacial es de primer orden, y resulta necesaria su consideración en todo momento.

Podemos ver más ejemplos de cómo la escala de análisis condiciona los resultados obtenidos. Supóngase un elemento lineal tal como un camino o el contorno de una finca cuyo perímetro quiere medirse. Como puede verse en la figura 11.3, la unidad de medida empleada provoca que se obtengan resultados distintos. Para medir la longitud de la línea utilizamos una unidad mínima, que podemos asimilar a una especie de «vara de medir». Todos los elementos de la línea que son menores que esa unidad mínima no se recogen. En el caso a) se obtiene un resultado de siete unidades. Si reducimos a la mitad la unidad, cabe esperar que la longitud sea el doble. Sin embargo, obtenemos un total de 17 unidades, de forma que la proporción entre el tamaño de nuestra vara de

medida y el número de unidades resultante no se mantiene.

Cuando esto sucede, podemos afirmar que carece de fundamento trabajar con una medida «absoluta» de longitud (u otro parámetro estudiado que se comporte de igual manera, tal como el perímetro de un área de estudio), y que esto solo tiene sentido dentro de un contexto dado que defina la forma en que se otienen los resultados. La unión de un valor resultante con la escala a la que se ha obtenido tiene en conjunto pleno significado, pero en casos como el anterior el valor resultante por sí mismo carece de dicho significado. Otra solución es la definición de parámetros invariantes a la escala, que no se ven afectados por esta.

El concepto de *fractal* tiene una implicación directa en este hecho. Para saber más sobre fractales, la referencia clásica es [86].

Por último, y para concluir este apartado, señalar que las implicaciones de la escala para el análisis se incorporan incluso en la representación y almacenamiento de los datos espaciales. Así, una ciudad puede definirse como un punto a una escala dada, y como un polígono si nos acercamos lo suficiente y estudiamos una porción concreta con más detalle. En función de su uso, puede ser más conveniente tratar el elemento *ciudad* de una u otra manera, lo cual también afecta al análisis del mismo.

En realidad, los conceptos *punto* y *línea* no existen como tales en el espacio geográfico. Un elemento tal como un cauce o una línea de alta tensión, que se recogen ambos en una capa vectorial como líneas, en realidad tiene un grosor. Lo mismo sucede con los elementos puntuales. Un árbol no es un punto, sino en todo caso un círculo. Por motivos de escala hacemos la abstracción de considerar puntos o líneas, porque a la escala habitual dichos elementos (árboles, caminos, etc.) pueden considerarse como tales.

También el propio formato de almacenamiento condiciona el efecto de la escala. Para el caso de datos vectoriales, existe el límite impuesto por la imposibilidad de almacenar números decimales de la precisión deseada. Es decir, la limitación del mínimo valor que puede almacenarse. No obstante, este límite es varios órdenes de magnitud inferior al definido por la precisión de los instrumentos de medida, con lo que no es considerable.

Una situación distinta es la que sucede con los datos ráster, donde el tamaño de celda está indirectamente condicionando una escala. La medición de áreas y distancias se encuentra influida por el tamaño elegido. Del mismo modo que no podemos recoger los detalles mínimos de una curva al utilizar una vara de medir de mayor tamaño, en el caso de una capa ráster, todo aquello que suceda en una escala inferior a la definida por el tamaño de celda queda ignorado. La resolución espacial es, por tanto, un elemento directamente relacionado con los resultados del análisis cuando se utilizan datos ráster.

11.2.2. El *Problema de la Unidad de Área Modificable*

Uno de los problemas principales asociados al análisis de datos espaciales es el relacionado con la definición de unidades de análisis. Muchas de las variables con las que trabajamos dentro de un SIG no pueden medirse de forma puntual, y por ello han de estudiarse para un área dada. Ejemplos de este tipo de variables son el porcentaje de población en un rango de edad determinado o la densidad media de población.

Las áreas que se definen para poder trabajar con las variables de esta índole son esencialmente arbitrarias. Por ejemplo, podemos estudiar el porcentaje de la población dentro de un intervalo de edad a nivel de país. La unidad *país* se establece sin ningún criterio propio del análisis

espacial, de igual modo que podría haberse realizado el mismo análisis a nivel de continente o de comarca, todas ellas divisiones por completo arbitrarias. No obstante, la utilización de una u otra unidad es problemática, ya que altera los resultados extraídos de las variables estudiadas.

Este problema, por tener relación con la elección de la unidad de agregación de la información, se conoce como *Problema de la Unidad de Área Modificable*(PUAM) [87] [1], y ha sido ampliamente estudiado en la literatura. Formalmente, puede definirse como «un problema causado por la imposición de unidades artificiales de definición espacial en fenómenos geográficos continuos, teniendo esto como consecuencia la generación de patrones artificiales» [88]. Aunque no se trata de una cuestión de reciente descubrimiento, la aparición de los SIG y las mayores capacidades de análisis que estos han propiciado ha atraído de nuevo el interés sobre el Problema de la Unidad de Área Modificable.

Los efectos del PUAM se pueden dividir en dos componentes: uno relacionado con la escala y otro relacionado con la agregación. El *efecto de escala* describe la variación de los resultados obtenidos en relación con el número de zonas en que se divide el total de la zona de estudio. Es decir, el tamaño de las unidades. Este efecto esta claramente relacionado con lo visto en el punto anterior.

Por su parte, el *efecto de zonificación* hace referencia a las diferencias que se producen cuando la información se agrega a una escala distinta. Por ejemplo, si se miden los datos de densidad de población por términos municipales, y posteriormente estos se agregan para presentarse por comunidades autónomas, ese cambio en la unidad de definición da lugar a diferencias en los valores resultantes.

Para darse cuenta de la importancia de este hecho, debe considerarse que una buena parte de la información geográfica que utilizamos en un SIG ha sido recogida originalmente a una escala distinta, y en ocasiones ha sufrido una agrupación en unidades mayores por motivos de mera facilidad de manejo.

Ambos efectos, el de zonificación y el de escala, no son independientes, sino que están íntimamente relacionados. La intensidad con que estos dos efectos afectan al análisis es variable, y existe asimismo una componente aleatoria. En líneas generales, el uso de unidades pequeñas implica que el número de elementos contenidos en las mismas es menor y por lo tanto estadísticamente menos fiable. En el extremo contrario, el uso de unidades grandes da valores estadísticamente más fiables pero oculta la variación que se produce dentro de las propias unidades.[89].

A pesar de tener una clara importancia en el análisis geográfico, las soluciones a la problemática que la definición de un área unitaria conlleva no son claras. Tradicionalmente se considera que se trata de un problema intratable. No obstante, algunos estudios [90] indican que existe un cierto grado de regularidad en los valores estadísticos agregados, dependiente de la autocorrelación espacial (ver siguiente punto) y la configuración de la variable.

Puede afirmarse que el Problema de la Unidad de Área Modificable es aún materia de amplio estudio, y el objeto de este estudio, que no es otro que el poder calcular los valores de los datos a la resolución espacial original (es decir, sin que los efectos de zonificación tengan relevancia), en caso de poder alcanzarse, requerirá un análisis sin duda complejo.

Un problema particular relacionado con el PUAM es la denominada *falacia ecológica*[87], que consiste en asumir que los valores calcula-

[1] *Modifiable Areal Unit Problem, MAUP*

dos para una unidad de área pueden aplicarse a los individuos de la población existente en dicha área. Sólo en el caso de que exista una completa homogeneidad para la variable analizada, lo cual muy raramente sucede, la anterior suposición sería cierta.

11.2.3. Autocorrelación espacial

Supóngase que se estudian una serie de poblaciones cercanas en las cuales se mide el porcentaje de personas afectadas por una determinada enfermedad infecciosa. Cabe esperar que, puesto que los habitantes de esas poblaciones están relacionados entre sí de diversas formas, la distribución de los valores recogidos obedezca en parte a la existencia de dichas relaciones. Por ejemplo, si en una población contraen la enfermedad un número dado de habitantes, es más factible que estos puedan contagiar a los de las poblaciones cercanas que a los de otros núcleos más alejados.

Por lo anterior, es probable que alrededor de una población con muchos casos de la enfermedad haya otras también con un elevado número de afectados, mientras que una población con pocos casos esté rodeada de otras también con escasa afección. Un comportamiento similar lo encontraríamos si midiéramos la concentración de un tóxico en distintos puntos de un embalse, ya que alrededor de un punto de alta concentración no parece lógico esperar concentraciones bajas.

Ejemplos como los anteriores cumplen lo que se conoce como *Primera Ley Geográfica de Tobler* [91], que establece que «todo está relacionado con todo, pero las cosas próximas entre sí están más relacionadas que las distantes».

De modo más formal, el termino *autocorrelación espacial* hace referencia a lo reflejado en los ejemplos anteriores, es decir, a la existencia de una correlación de la variable consigo misma, de tal modo que los valores de esta variable en un punto guardan relación directa con los de esa misma variable en otros puntos cercanos.

En el caso de la enfermedad infecciosa o la concentración del producto tóxico, los valores altos suelen tener en su entorno valores también altos, y de modo similar sucede para valores bajos. Se dice que existe una *autocorrelación espacial positiva*. Puede, no obstante, existir una *autocorrelación espacial negativa*, si los valores altos se rodean de valores bajos y viceversa.

En caso de no existir ningún tipo de autocorrelación espacial, se tiene que los datos recogidos en una serie de puntos son independientes entre sí y no se afectan mutuamente, si que tenga influencia de la distancia.

La figura 11.4 muestra unas sencillas capas ráster en las que se presentan los tres tipos de autocorrelación espacial anteriores.

Las consecuencias de la existencia de autocorrelación espacial son numerosas y de gran importancia.

Por una parte, muchos de los análisis estadísticos suponen la independencia de la variable. Puesto que existe una dependencia de la componente espacial, será necesario para obtener resultados correctos introducir dicha componente espacial como una variable más.

Existiendo autocorrelación espacial, y siendo esta positiva, la inferencia estadística es menos eficaz que si se cuenta con un número igual de observaciones de una variable independiente. Es decir, se pierde parte de la capacidad explicativa de los datos. Esto se materializa en mayores varianzas en las estimaciones y peores ajustes de modelos, entre otras consecuencias.

A pesar de ello, puede sacarse también provecho de la existencia de una dependencia espacial. Puesto que los puntos cercanos a uno dado guardan relación con este, pueden emplearse para estimar su valor, siendo este el fundamento prin-

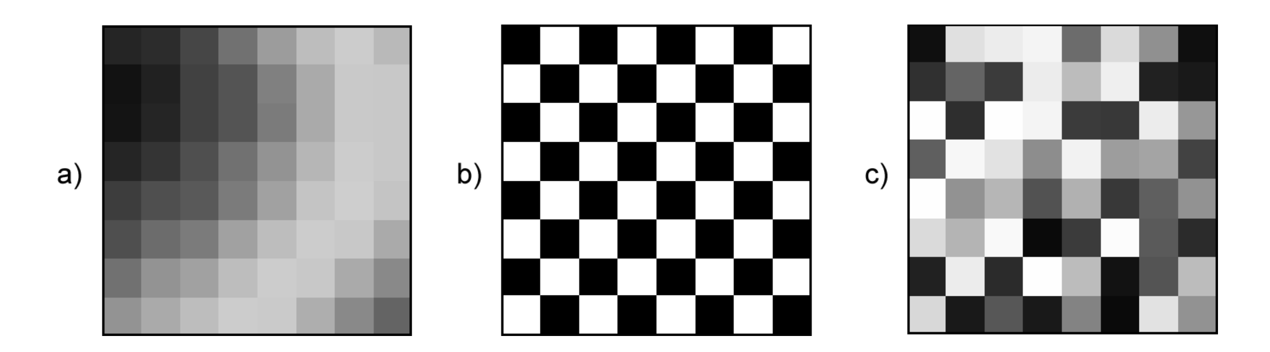


Figura 11.4: a) Autocorrelación espacial positiva. b) Autocorrelación espacial negativa. c) Ausencia de autocorrelación espacial (independencia)

cipal de los distintos métodos de interpolación (Capítulo 14).

En lugar de incorporar la autocorrelación espacial como un elemento más, otra forma de proceder es analizar la intensidad de esta para ver en qué medida lo anterior es cierto o no. Así, el estudio de la autocorrelación espacial puede servir para juzgar si procede la aplicación de métodos estadísticos que no consideren la dependencia espacial. Como veremos en el capítulo 13, si a través de los valores de los indicadores correspondientes podemos aceptar la hipótesis nula de ausencia de dependencia espacial, entonces los inconvenientes anteriormente citados pueden no existir.

Como ya venimos observando, el conjunto de conceptos básicos sobre datos espaciales que estamos viendo en esta sección no es un conjunto de elementos independientes. Por ejemplo, la autocorrelación espacial se halla directamente ligada con el concepto de escala, y un cambio de escala puede hacer que la autocorrelación cambie de signo [92]. Veamos un ejemplo.

Sea un monte en el que los árboles grandes están separados una distancia dada por el efecto de la competencia, y entre los cuales crecen los árboles más pequeños. Supongamos que la distancia media entre árboles grandes es de unos 20 metros. Si hacemos un muestreo en el que medimos la altura media de los árboles en parcelas separadas aproximadamente cada 10 metros, es probable que midamos alternamente una parcela con un árbol grande y una con algunos pequeños, de forma que tendremos una marcada autocorrelación espacial negativa. Si por el contrario medimos parcelas de un metro de radio separadas a su vez un metro, mediremos muchas parcelas cercanas en las que solo entrarán árboles pequeños que se agrupan bajo los grandes, de tal forma que la autocorrelación espacial que obtendremos será positiva.

Es importante considerar todos estos factores de forma global, pues todos ellos tienen importancia y afectan al trabajo con datos geográficos.

11.2.4. Existencia de estructura

Tanto la disposición de los datos como las propiedades de la variable estudiada (por ejemplo, la propia autocorrelación espacial como propiedad intrínseca), exhiben una estructura determinada. En la figura 11.5 pueden verse dos conjuntos de puntos distintos, sobre los cuales cabe plantearse si los resultados obtenidos de su análisis pueden darse como igual de fiables. Puesto que

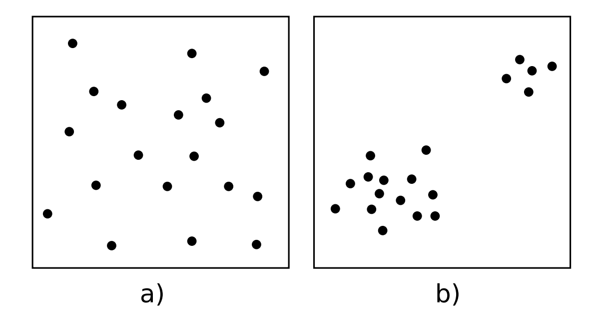


Figura 11.5: Dos estructuras distintas con diferentes implicaciones a la hora del análisis de los datos que representan

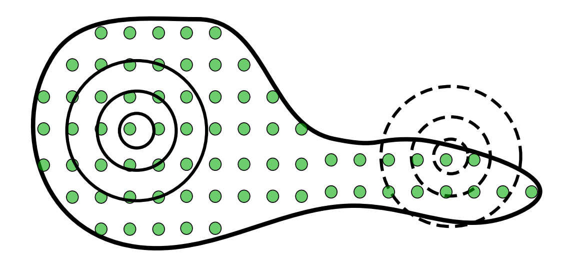


Figura 11.6: Representación del efecto borde y cómo este afecta en mayor o menor medida en función de la escala de análisis. Las zonas en trazo continuo no se ven afectadas. Las zonas en trazo punteado están afectadas de efecto de borde en diferente grado.

la estructura espacial de ambos es distinta y la componente espacial juega un papel importante, esta estructura puede condicionar los resultados y tener influencia sobre ellos.

Por ejemplo, vemos que en el patrón b) los puntos se hallan más agrupados, mientras que en el a) los puntos están distribuidos uniformemente a lo largo de la extensión de la zona de análisis. Si existe autocorrelación espacial positiva, la información recogida en el patrón b) es mucho menos representativa, ya que los puntos cercanos recogen información en cierta medida redundante. A pesar de disponer de un numero n de valores recogidos en otros tantos puntos, el análisis estadístico de estos no es tan preciso como si se dispusiera de n observaciones independientes. En realidad, los resultados que obtendremos serán como si hubiéramos muestreado un número menor de puntos que los que realmente tenemos.

Los dos principales conceptos estadísticos que definen la estructura espacial de los datos son la *estacionaridad* y la *isotropía*. Estos se estudian principalmente en relación a los denominados efectos de primer y de segundo orden. El efecto de primer orden es el valor esperado, es decir, la media. El de segundo orden es la covarianza entre distintas zonas.

La estacionaridad indica que el proceso es invariante a la traslación. Es decir, que las propiedades son constantes en el espacio y no existe tendencia alguna. La isotropía indica que el proceso es invariante a la rotación. Un proceso cuyas propiedades de segundo orden son isotrópicas es aquel en el que la covarianza presenta la misma variación en todas direcciones.

Veremos en diversos puntos de esta parte del libro como la presencia de isotropía o su ausencia (anisotropía) tiene importancia a la hora de realizar distintos tipos de análisis.

11.2.5. Existencia de tendencias espaciales

Podemos decir que existe una tendencia espacial cuando los valores de una variable están relacionados con sus propias coordenadas geográficas. Por ejemplo, existe una tendencia a que la temperatura disminuya conforme nos alejamos del ecuador. Por ello, en un mapa de temperaturas para una región lo suficientemente amplia, cabe esperar valores menores en el extremo más distante del ecuador.

El dato de localización geográfica plantea un contexto dentro del cual se sitúan los restantes valores, en este caso, la temperatura observada.

Esto hace que el mismo valor de una variable no tenga el mismo significado cuando aparece en un punto que cuando lo hace en otro. No es lo mismo un valor de temperatura de 40°C en Madrid que en Oslo. El valor en sí es idéntico, pero su interpretación es distinta.

Conocer las tendencias existentes para una variable nos ayuda a comprender mejor esta y analizarla de forma correcta. Si es posible cuantificar dicha tendencia, resulta factible eliminar su influencia de los datos, de forma que estos ya no se vean afectados por ella, o bien considerarla explícitamente como parte del análisis.

Las consecuencias de la existencia de tendencias son similares a las que se derivan de la presencia de autocorrelación espacial, ya que invalidan el supuesto de independencia de los datos.

11.2.6. Efectos de borde

Las zonas que estudiamos dentro de todo análisis espacial tienen unos límites establecidos. Estos límites vienen definidos de forma artificial —el límite de la fotografía aérea de la que disponemos, por ejemplo— o bien de forma natural —si estudiamos un bosque junto a un pantano, el bosque encuentra su límite al borde de este último.

Imaginemos un caso como este segundo y observemos la figura 11.6.

Si dentro del bosque los árboles están plantados de forma regular (supongamos que es una repoblación con un marco fijo), se puede decir que en cualquier punto dentro de esa masa existe una densidad constante. En otras palabras, si nos situamos en cualquier punto de dicha masa, ya sea cerca o lejos del borde, los árboles están plantados con una misma densidad. No obstante, para el cálculo de la densidad necesitamos establecer un área de análisis, puesto que no es una variable que pueda computarse puntualmente. Debido a esto, la densidad no será constante al medirla en todos los puntos, ya que en las zonas de borde existirá una parte de dicho área que se encontrará fuera de la masa de bosque, con lo que el número de pies será menor (no hay árboles en la zona limítrofe, es decir, el embalse), y por tanto también lo será la densidad.

El efecto de borde no es independiente de otros elementos como la escala, ya que la escala de análisis tiene un influencia directa en él. Como se ve en la propia figura 11.6, el porcentaje del círculo de análisis que queda fuera de la zona de bosque es distinto en función del tamaño de dicho círculo.

Otros análisis que en breve veremos hacen uso de un mecanismo similar. Por ejemplo, analizando el número de puntos situados a una distancia menor que un umbral dado. En los puntos cerca del borde, la presencia de dicho borde va a distorsionar los valores calculados. Como también veremos, las distintas formulaciones tienen en muchos casos expresiones corregidas que modifican los valores obtenidos en función de la distancia al borde.

En general, es importante considerar los efectos de borde para saber si los valores calculados dentro de cualquier análisis estadístico son válidos o no. Cuando nos encontramos lo suficientemente cerca de un borde (sea este uno artificial como el borde de la capa o uno natural dentro de la propia capa tal como el mencionado límite de un bosque), la información que derivamos de los datos espaciales puede ser incoherente con la realidad.

Veremos ejemplos variados a lo largo de los siguientes capítulos, no solo relacionados con el análisis de datos puntuales como en los casos comentados anteriormente. En el apartado 15.2.2 veremos cómo el efecto de borde afecta a un tipo particular de análisis sobre capas ráster. En otros casos, el efecto de borde no se manifiesta única-

mente para puntos cercanos a dicho borde, sino para todos aquellos relacionados o conectados con él, con independencia de su distancia a este. Veremos este caso en el apartado 16.6.3.

Con relación a este último supuesto, no debe olvidarse nunca que los procesos que estudiamos y que analizamos a través de la información espacial están influenciados por otros procesos que pueden darse fuera del marco delimitado sobre el que trabajamos, alejados de él e incluso a una escala distinta. Así, estudiar la vegetación de una zona dada implica estudiar el clima que la condiciona. Aunque el relieve y las condiciones locales son los que afectan a este en primera instancia, el clima es un proceso global que opera a una escala mayor a la de la zona cuya vegetación estudiamos, y efectos fuera de dicha zona pueden tener repercusión sobre ella.

11.2.7. Localización representada

Como veíamos al tratar el Problema del de Unidad de Área Modificable, algunas de las variables geográficas requieren un área para ser recogidas, y no pueden hacerse de forma puntual. En otros casos, la necesidad de establecer unidades no puntuales no viene motivada por la variable recogida o la estructura geográfica que se estudia, sino por la forma de almacenar la información de dicha variable. Tal es el caso del modelo ráster, en el que el territorio se divide en unidades geométricas arbitrarias, generalmente unidades regulares de forma cuadrada.

Para cada una de estas unidades, se tiene un valor de la variable estudiada, pero lo que dicho valor representa en el territorio puede variar en función del criterio establecido. Como se recoge en la figura 11.7, en la cual la variable recogida es la elevación, el valor de cada celda puede ser la elevación en el centro de la celda o bien el valor medio de toda ella, entre otras opciones posibles.

Este tipo de cuestiones deben considerarse al trabajar con los datos espaciales, y homogeneizar los criterios en la medida de lo posible, siempre considerando la naturaleza de la variable recogida.

11.3. Algunos cálculos espaciales básicos

La mayor parte de los análisis espaciales hacen uso de cálculos geométricos sencillos, a partir de los cuales se construyen algoritmos más complejos. Veremos en esta sección esos cálculos básicos, que constituyen los fundamentos del análisis geométrico tanto en el plano como en el espacio.

La idea de distancia es fundamental para todo análisis espacial. En el plano, la distancia euclídea entre dos puntos dados es

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \qquad (11.3.1)$$

En el análisis geográfico es habitual utilizar la denominada *distancia de Manhattan*[2], cuya expresión es

$$d_m = (x_2 - x_1) + (y_2 - y_1) \qquad (11.3.2)$$

Tanto la distancia euclídea como la de Manhattan son casos particulares de las denominadas *métricas LP*, que responden a una expresión de la forma

$$d^\beta = (\|x_2 - x_1\|^p + \|y_2 - y_1\|^p)^{\frac{\beta}{p}} \qquad (11.3.3)$$

En el caso de ser $p = 1$ se tiene la distancia de Manhattan, y para $p = 2$ la distancia euclídea.

[2]Se denomina así debido a que es similar a la recorrida por las calles regularmente dispuestas tales como las de la ciudad de Manhattan.

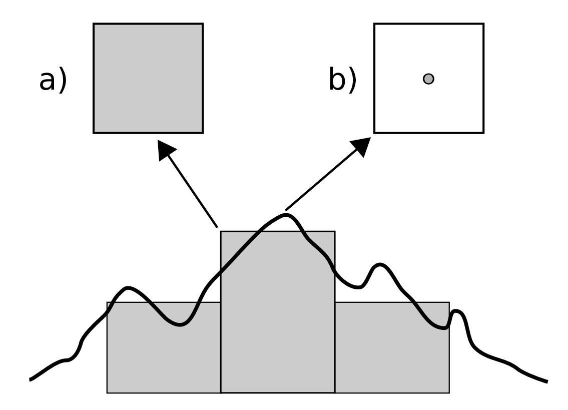


Figura 11.7: El valor recogido en una unidad puede interpretarse con distintos criterios. a) Media de la celda. b) Valor en el punto medio.

Cuando se utilizan capas ráster, el concepto de distancia puede entenderse de un modo distinto. Como resulta lógico, puede aplicarse la distancia euclídea entre los centros de las celdas, pero en ciertos casos puede ser conveniente trabajar no en coordenadas geográficas, sino de celdas, ya que, como sabemos, el espacio se divide en un número finito de estas en una capa ráster. Por esta razón, y puesto que las coordenadas de celda son expresadas en números enteros de la forma (fila, columna), resulta además conveniente que esa distancia sea también un valor entero[93].

Sobre este planteamiento pueden definirse distintos tipos de distancia ráster considerando principalmente el número de celdas por las que debe pasarse para ir de una celda a otra. Por ejemplo, si se permite el movimiento en todas direcciones, la distancia desde una celda a las ocho que la rodean es igual a 1 en todos casos, pues se realiza en un único paso. Por similitud a la forma en que uno puede moverse en un tablero de ajedrez, este tipo de distancia se conoce como distancia *de tablero de ajedrez*[3].

Si, por el contrario, se permite tan solo el movimiento en dirección vertical y horizontal, la distancia a las celdas diagonales —por ejemplo, desde la celda (x, y) hasta la $(x + 1, y + 1)$— es igual a 2. En este caso tenemos la anteriormente mencionada distancia de Manhattan.

En la figura 11.8 pueden verse los valores de distancia entre una celda central y sus circundantes según las definiciones de distancia anteriores, junto con otras como la distancia *ortogonal* o la distancia *Chamfer 3–4*[94]. El objetivo de estas distancias es mitigar en cierta medida la distorsión que se produce con las otras distancias ráster a medida que aumenta el alejamiento.

El análisis de costes se lleva a cabo en un SIG esencialmente en formato ráster, por lo que lo anterior es de importancia al respecto, y se detallará en el capítulo 20.

Además de hallarse las distancias entre puntos concretos, pueden calcularse entre geometrías. La distancia entre dos rectas en el plano es igual a la distancia entre un punto cualquiera de una de ellas a la otra en el caso de que sean rectas paralelas. Si no lo son, la distancia es nula, ya que existirá un punto en el que se corten. No obstante, no ha de olvidarse que en un SIG habitualmente no trabajamos con rectas de longitud infinita en

[3]Chessboard distance

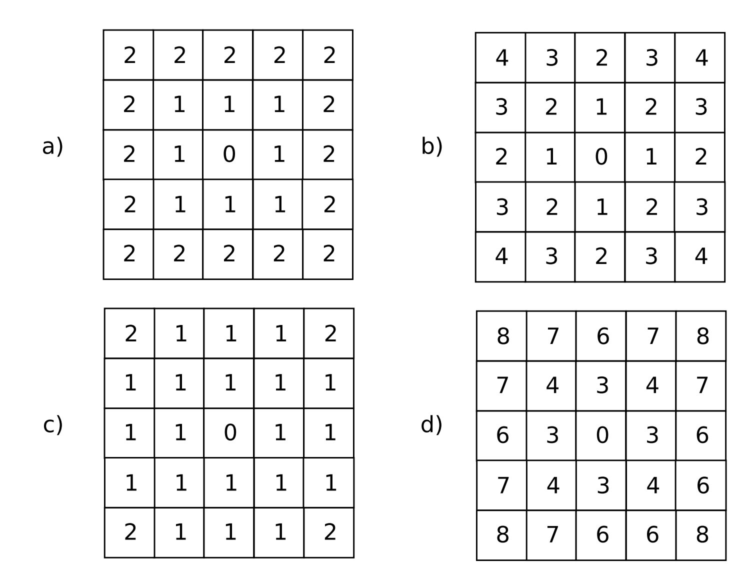


Figura 11.8: Distintos tipos de distancia ráster: a) tablero de ajedrez, b) Manhattan, c) ortogonal, d) Chamfer 3–4

el sentido matemático, sino con segmentos de estas.

La distancia de un segmento definido por sus extremos (x_1, y_1) y (x_2, y_2) a un punto de coordenadas (x_3, y_3) se calcula como la distancia de este último hasta la intersección de la recta que pasa por el mismo y es perpendicular al segmento. Dicho punto de intersección tiene por coordenadas

$$x = x_1 + u(x_2 - x_1) y = y_1 + u(y_2 - y_1) \quad (11.3.4)$$

donde u se calcula según

$$u = \frac{(x_3 - x_1)(x_2 - x_1) + (y_3 - y_1)(y_2 - y_1)}{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (11.3.5)$$

.

La distancia entre un punto y un polígono es la de dicho punto a la línea que contiene al segmento más cercano de cuantos componen el perímetro del polígono.

Para el caso de polígonos, dos son las magnitudes principales: área y perímetro. El área se calcula aplicando la fórmula

$$A = \left| \frac{1}{2} \sum_{i=1}^{n} x_i y_{i+1} - x_{i+1} y_i \right| \quad (11.3.6)$$

donde se considera que el vértice $n + 1$ se corresponde con el primero, esto es, el polígono es una polilínea cerrada.

Para aquellos polígonos que contengan «huecos», basta restar del área total la correspondiente a esos huecos. Esta última se calcula de igual modo, ya que los huecos están definidos de forma similar por un conjunto de puntos conectados.

El perímetro de un polígono es la suma de las distancias entre vértices consecutivos, es decir,

$$P = \sum_{i=1}^{n} \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2} \quad (11.3.7)$$

Además de los anteriores, un parámetro de interés también para polígonos es el centro de gravedad, cuyas coordenadas se calculan según

$$C_x = \frac{1}{6A}\sum_{i=1}^{n}(x_i x_{i+1})(x_i y_{i+1} - x_{i+1} y_i)$$
$$C_y = \frac{1}{6A}\sum_{i=1}^{n}(y_i y_{i+1})(x_i y_{i+1} - x_{i+1} y_i) \qquad (11.3.8)$$

La medida del área y de la longitud de un elemento lineal como el perímetro de un polígono o una recta, pueden llevarse a cabo para datos en formato ráster de una forma distinta. Para el caso del área basta contar el número de celdas del polígono y multiplicarlo por el área de una única celda. En el caso de la longitud, basta sumar la longitud total de todos los lados exteriores, esto es, de aquellos que no son contiguos a otra celda del polígono. Todos estos cálculos se establecen en función del tamaño de celda como magnitud base. Para el cálculo del centroide, este es el centro de masas calculado como si cada celda perteneciente al polígono fuese una masa puntual unitaria.

Para concluir, un sencillo análisis entre un punto y un polígono, el cual utilizaremos frecuentemente, es la comprobación de si este punto se encuentra dentro o fuera del polígono. Para ello existen diversas metodologías, pero la más habitual es la basada en el número de veces que una semirecta con origen en el punto cruza el borde del polígono. El algoritmo es como sigue [95]:

- Se traza una recta desde el punto en cuestión hasta un punto fuera del polígono. Lo habitual es considerar la semirecta horizontal desde el punto dado y bien en la dirección positiva o bien en la negativa.
- Se cuenta el número de veces que dicha semirecta corta la frontera del polígono.
- Si el número de cortes es par, el punto se encuentra fuera. Si es impar, el punto se encuentra dentro.

En la figura 11.9 se muestra un ejemplo de lo anterior.

La pertenencia o no del punto al polígono queda definida así en todos los casos, salvo cuando el punto está en la propia frontera o bien la semirecta coincide en algún tramo con el contorno, en cuyo caso resulta imposible el cálculo del número de cortes (Figura 11.10).

11.4. Relaciones espaciales

Como ya sabemos, conceptos tales como la posición o el tamaño son básicos para el análisis geográfico, pues derivan de la propia georreferenciación inherente a todo dato espacial. El hecho de que exista dicha referencia en el espacio es responsable de que los mismos valores de una variable no tengan igual significación en unos lugares que en otros, y que estos lugares no solo se consideren en términos absolutos, sino también relativos entre los distintos datos espaciales.

La importancia de esta posición relativa ya la vimos al tratar la autocorrelación espacial, ya que una misma serie de valores, si se disponen de una forma distinta, pueden presentar un signo distinto de autocorrelación espacial, con las consecuencias que ello tiene.

Si pensamos por ejemplo en el uso de otro tipo de información geográfica tal como la de un callejero urbano para orientarnos en una ciudad, utilizamos ideas tales como «la Calle Mayor *es paralela* a esta avenida» o «El teatro al que me dirijo está *detrás* de ese bloque de edificios». Existe de igual modo una relación entre los distintos elementos, que es la que permite que podamos analizar y explotar la información geográfica, pues esta en gran medida no tiene sentido como una colección de datos aislados.

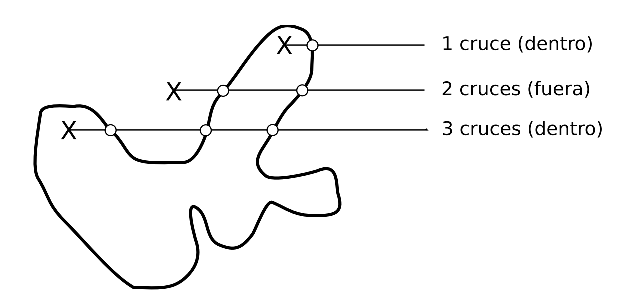


Figura 11.9: Pertenencia de un punto al interior de un polígono en función del numero de cortes entre la frontera de dicho polígono y una semirecta con extremo en dicho punto.

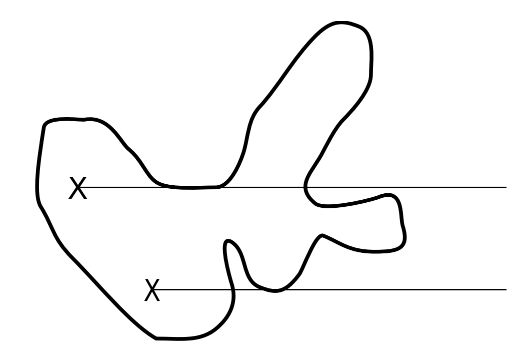


Figura 11.10: Problemas de la metodología para determinar si un punto se encuentra en el interior de un polígono cuando la semirecta coincide parcialmente con la frontera.

Así pues, resulta claro que los distintos elementos con los que trabajamos dentro de una o varias capas de información geográfica se relacionan entre sí. Estas relaciones pueden obedecer a diversos criterios y son la base de un gran número de distintos procedimientos que las estudian y generan resultados en función de ellas.

De entre dichas relaciones, algunas son de tipo topológico y otras se fundamentan no en la topología existente, pero sí en otras propiedades de tipo espacial, por ejemplo propiedades métricas como la distancia. Además de lo anterior, existen muchos otros criterios en base a los cuales pueden clasificarse las relaciones.

En esta sección daremos una definición formal de los principales tipos de relaciones y, especialmente, de los razonamientos que dan lugar a estos criterios y son claves para comenzar a entender el análisis espacial tal y como este se presenta en un SIG. De esta forma, posteriormente podremos aplicar estas relaciones con claridad a los distintos datos geográficos.

[96] propone los siguientes tipos de relaciones espaciales:

- Relaciones direccionales, que describen el orden en el espacio. Por ejemplo, *al norte de*, *al sur de*, etc.
- Relaciones topológicas, las cuales describen la vecindad e incidencia. Por ejemplo, *son disjuntos* o *son adyacentes*.
- Relaciones comparativas, que describen la inclusión. Por ejemplo *está en*.
- Relaciones de distancia, tales como *lejos de* o *cerca de*.
- Relaciones «difusas» tales como *al lado de* o *a continuación*.

Las relaciones espaciales pueden establecerse entre todas las combinaciones posibles de entidades geográficas. Por nombrar algunos ejemplos, las siguientes cuestiones se refieren a relaciones entre objetos geográficos de diversa índole:

- ¿Se encuentra esta localización a menos de 100 metros en línea recta de algún camino? (relación entre un punto y una recta)
- ¿Cruza ese camino algún área protegida? (relación entre una recta y un polígono)
- ¿Cruza ese camino bajo alguna línea de alta tensión? (relación entre dos líneas)
- ¿Existe algún área urbanizada contigua a ese área protegida? (relación entre dos polígonos)

Las relaciones pueden establecerse entre elementos con un mismo tipo de información, o bien entre tipos distintos. Los anteriores son ejemplos de este último caso. Un ejemplo del primero podría ser la relación de proximidad entre dos emplazamiento puntuales de una misma clase (¿existe una farmacia a menos de un kilómetro de esta otra farmacia?).

Dentro de un SIG, las relaciones topológicas tienen utilidad en los procesos de análisis implementados como tales, pero también en otras partes de un SIG que, constituyendo análisis propiamente dichos, quizás no se perciben como tales. Por ejemplo, las herramientas de selección de entidades dependen de las relaciones espaciales que estas presentan con el objeto empleado como criterio de selección, ya sea este un punto concreto que el usuario escoge con el ratón, un área rectangular delimitada de igual modo, o las entidades de otra capa adicional, entre otros.

A la hora de clasificar y definir las relaciones espaciales, deben considerarse tres enfoques principales: un enfoque netamente matemático, un enfoque psicológico y un enfoque geográfico. El enfoque matemático pretende formalizar las distintas relaciones, de manera que puedan estudiarse y analizarse a través de las herramientas matemáticas habituales, tanto topológicas como espaciales. Por su parte, el enfoque geográfico surge según se desarrollan los Sistemas de Información Geográfica y aparece la necesidad de expresar las relaciones espaciales de un modo adecuado para implementar estas, así como los distintos algoritmos que se sustentan en ellas. Puede entenderse en cierta forma como una versión práctica del enfoque matemático.

Tanto el enfoque matemático como el geográfico son netamente cuantitativos pero a la hora de comunicar algún tipo de conocimiento espacial que lleve implícita una relación espacial, lo hacemos principalmente de forma cualitativa [97] [98].

Así, al indicar a otra persona si se puede llegar rápidamente a una dirección dada dentro de la ciudad, no decimos «el parque al que quieres ir está contenido dentro de un radio de 1,2 km» sino que diremos algo como «sí, está cerca, se puede llegar andando». En nuestro pensamiento espacial y en el lenguaje que utilizamos para expresarlo, no es necesaria la precisión cuantitativa, que sin embargo sí se requiere para plantear otros modelos de relaciones. Entender las relaciones espaciales cualitativas para poder implementarlas en una herramienta lógica como un SIG es en esencia un problema de traducción entre un lenguaje natural y uno formal [99].

La forma en que los SIG incluyen las relaciones espaciales para sus propósitos debe combinar todos estos enfoques con objeto de conseguir que el razonamiento espacial pueda transmitirse de forma sencilla y lo más efectiva posible. Teniendo en cuenta esto, autores como [100] argumentan que, en la actualidad, la falta de un sistema de relaciones espaciales completo que dé respuesta a todas las necesidades que se plantean es uno de los principales escollos para un mayor desarrollo de la disciplina de los SIG. El problema, no obstante, no presenta una solución sencilla, ya que, como hemos visto, los criterios a aplicar pueden ser muy variados y las ideas matemáticas han de combinarse igualmente con los elementos perceptivos acerca de cómo estas relaciones se entienden y se interpretan [101].

Lo habitual dentro de un SIG es la conversión de los conceptos del lenguaje natural (cualitativos) en elementos cuantitativos, de forma que estos pueden después tratarse con las herramientas de algún sistema formal de relaciones.

Es importante reseñar que las relaciones geográficas, sea cual sea el criterio por el que se definan, no están condicionadas de forma alguna al

tipo de almacenamiento del dato espacial (vectorial, ráster, etc) u otras características arbitrarias del mismo. Son, por el contrario, conceptos puramente teóricos sobre elementos situados en el espacio, los cuales pueden aplicarse a cualquier objeto con independencia de cómo este haya sido recogido. Aun así, la forma de almacenamiento condiciona en cierta medida las relaciones existentes o, al menos, la forma en que estas relaciones se incluyen en el propio almacenamiento. Así, para el caso por ejemplo de una capa ráster, tenemos una estructura regular de elementos relacionados entre sí de tal forma que son contiguos y están a una misma distancia. Es decir, con una relación topológica y otra métrica que se mantienen constantes para todos los elementos unitarios mediante los cuales se almacena la capa.

11.4.1. Relaciones topológicas

Entrando en la propia definición de relaciones, el conjunto principal de estas es el formado por las de tipo topológico, que serán por ejemplo las que empleemos para combinar las geometrías y elementos de dos capas vectoriales según cómo sean dichas relaciones entre ellas. De entre estas relaciones destacan los denominados *predicados espaciales*, que son operaciones de tipo lógico que nos indican si entre dos objetos geográficos existe o no un tipo de relación dada. Se consideran estos objetos en $\mathbb{R}^2$, es decir, como objetos planos.

La definición formal de estos predicados ha sido motivo de abundante estudio desde la aparición de los SIG, en parte motivado por la mayor necesidad que se tiene de tal formalismo si se pretende estructurar adecuadamente todas las operaciones de análisis que un SIG puede contener.

Uno de los sistemas iniciales de predicados es el conocido como *4–Intersection* [102]. Según este modelo, la relación entre dos objetos A y B queda definida por las intersecciones entre las fronteras (δA y δB) y los interiores (A y B) de estos. Se tienen así cuatro intersecciones con las que se conforma una matriz que caracteriza la relación existente.

$$\Im_4(A,B) = \begin{pmatrix} A \cap B & A \cap \delta B \\ \delta A \cap B & \delta A \cap \delta B \end{pmatrix} \quad (11.4.1)$$

Para cada una de las cuatro intersecciones se estudia algún invariante topológico, es decir, alguna propiedad que sea invariante a las transformaciones topológicas. De entre ellas, lo más habitual es emplear el contenido, esto es, si la región delimitada por la intersección esta vacía ($\varnothing$) o no ($\neg\varnothing$).

Teniendo cuatro elementos y dos posibles valores para cada uno, existen un total de $2^4 = 16$ diferentes matrices con la forma anterior. De estas, ocho pueden darse en un plano entre objetos bidimensionales con fronteras cerradas, cada uno de los cuales define una *región*.

Un razonamiento similar puede aplicarse al caso de líneas, cuya principal diferencia radica en que conforman elementos con fronteras no cerradas. La forma de proceder y las relaciones definidas son análogas en gran medida.

A partir del modelo *4–Intersection*, Egenhofer [102] desarrolla el modelo *9–Intersection*, en el cuál se amplia el anterior a la consideración de tres elementos en lugar de dos. Además de considerar las fronteras e interiores de los objetos A y B, se consideran asimismo los exteriores de los mismos (A^- y B^-). La matriz característica queda entonces de la forma

$$\Im_9(A,B) = \begin{pmatrix} A \cap B & A \cap \delta B & A \cap B^- \\ \delta A \cap B & \delta A \cap \delta B & \delta A \cap B^- \\ A^- \cap B & A^- \cap \delta B & A^- \cap B^- \end{pmatrix} \quad (11.4.2)$$

El número total de matrices posibles es en este caso de $2^9 = 512$. De todas ellas, solo un pequeño subconjunto representan relaciones posibles en $\mathbb{R}^2$ a las cuales pueda asignarse una interpretación geométrica.

Por ejemplo, la matriz siguiente, en la que todos los elementos son el conjunto vacío, resulta imposible de obtener con ningún tipo de relación.

$$\Im_9(A,B) = \begin{pmatrix} \emptyset & \emptyset & \emptyset \\ \emptyset & \emptyset & \emptyset \\ \emptyset & \emptyset & \emptyset \end{pmatrix} \tag{11.4.3}$$

Dependiendo del tipo de objetos sobre el que se den las relaciones, el modelo *9–Intersection* amplía al *4–Intersection* de una u otra forma.

En el caso de dos regiones, se tienen ocho posibles relaciones, por lo cual no existe diferencia entre ambos modelos.

Para el caso de dos líneas en $\mathbb{R}^2$, aparecen 25 nuevas relaciones. En caso de considerar líneas ramificadas (con más de dos puntos extremos), aparecen además 21 relaciones adicionales. Por último, para el caso de una línea y una región, se tienen un total de 19 relaciones posibles, 20 en el caso de admitirse líneas ramificadas.

11.4.2. Índices métricos

Pese a su aparente complejidad y completitud, el modelo *9–Intersection* deja de lado otra serie de relaciones posibles, tales como las basadas en distancias u orientaciones, las cuales son en muchos casos más cercanas al habla común y al enfoque perceptivo y lingüístico del razonamiento espacial. Estas relaciones pueden definirse formalmente no a través de predicados como los establecidos por los modelos anteriores, sino cuantificándose mediante índices diversos. El uso de estos índices enriquece la definición de las distintas relaciones expresadas mediante un modelo como el *9–Intersection*, añadiendo información acerca de la naturaleza exacta de estas.

Por ejemplo, si dos regiones de una hectárea se intersecan, no es lo mismo que lo hagan dando lugar a una intersección de media hectárea que a una de 100 metros cuadrados. Topológicamente, se trata de la misma relación, pero está claro que, en la práctica, las implicaciones de una u otra intersección son bien distintas.

Dependiendo de los tipos de entidades que se consideren, existen distintos índices que cuantifican la relación existente. [103] propone para el caso de una región y una línea el análisis en términos métricos de las siguiente propiedades:

- Subdivisión. Se definen índices que describen la forma en que la frontera, interior y exterior de la región subdivide a la frontera y el interior de la línea. Estos índices tratan, entre otros aspectos, la forma en que la línea divide el interior de la región, el exterior de esta (pudiendo generar áreas delimitadas por la línea y la región en el exterior de esta última), la relación entre la frontera de la región y la línea, o cómo el perímetro de la región puede quedar dividido en distintos tramos por las intersecciones con la línea.
  Por ejemplo, la *relación de subdivisión del área interior* (*internal areasplitting ratio(IAR)*) se define cómo el mínimo área de las dos que quedan a cada uno de los lados de la línea dentro de la región, dividido por el área total de región.

  $$IAR = \frac{a_{min}}{a_{total}} \tag{11.4.4}$$

  Para una descripción más detallada de otros índices puede consultarse la referencia original.
- Cercanía. Los índices de cercanía cuantifican el alejamiento entre partes disjuntas de los objetos relacionados. Para su cálculo, se utilizan medidas de distancia como las descritas en 11.3. Cuatro son los índices definidos, que miden

- La distancia entre la frontera de la línea y la de la región, cuando la línea está en el exterior de la región.
- La distancia entre la frontera de la línea y la de la región, cuando la línea está en el interior de la región.
- La distancia del recorrido mínimo entre el interior de la línea y la frontera de la región si el interior de la línea está en el exterior de la región.
- La distancia del recorrido mínimo entre el interior de la línea y la frontera de la región si el interior de la línea está en el interior de la región.

Para el caso de dos líneas, [104] propone estudiar también las mismas propiedades —subdivisión y cercanía—, desarrollando un planteamiento similar. [98], por su parte, añade elementos direccionales a las relaciones entre líneas, definiendo un ángulo local (el ángulo puntual en el punto de corte) y uno global (el definido por las direcciones globales de las líneas). Asimismo, incluye relaciones entre los rectángulos mínimos que engloban a las líneas, teniendo de este modo relaciones de área que complementan a las anteriores.

11.4.3. Otras relaciones

Muchas otras relaciones se pueden establecer entre elementos espaciales, si bien las anteriores son las principales y las que se presentan como más adecuadas para formalizar los análisis que dependen de ellas. Otros análisis que veremos más adelante implican relaciones espaciales basadas en otra serie de conceptos.

Por ejemplo, el análisis hidrológico implica el estudio de la conectividad hidrológica entre sus elementos. Estos pueden ser celdas en una capa ráster o triángulos en un TIN, entre otros, y en función de los valores asociados a ellos, en particular la elevación, se establecen las relaciones de conectividad. Junto a las expresiones *cerca, lejos, junto a, a la derecha* u otras tantas que ya hemos visto para las relaciones métricas o topológicas, podemos emplear otras asociadas a estas relaciones de conectividad y decir, por ejemplo, que «el pueblo se encuentra *aguas arriba* de la presa».

De un modo similar, los análisis de visibilidad establecen una relación entre los elementos, según estos puedan verse entre ellos o no, y el análisis de una serie de puntos situados sobre una red también implica una conectividad.

Las relaciones de este tipo no conforman sistemas completos formales como las relaciones topológicas que se han desarrollado anteriormente, pero su importancia para estudios particulares debe considerarse y conocerse, entendiendo que se tratan igualmente de relaciones basadas en la posición espacial de los elementos.

11.5. Resumen

Los datos espaciales presentan particularidades que tienen una gran importancia en los procesos de análisis. Entre estas, la existencia de una estructura, la presencia de efectos de borde o los efectos de escala y derivados tales como el denominado Problema de la Unidad de Área Modificable, son los más relevantes.

La autocorrelación espacial es otro de los elementos que siempre deben tenerse en cuenta a la hora de estudiar los datos espaciales, pues condiciona los resultados de los análisis según sea dicha autocorrelación.

Además de lo anterior, los distintos elementos con los que trabajamos en el análisis espacial se relacionan entre sí. El estudio y clasificación de dichas relaciones presenta alternativas diversas que tratan de recoger la totalidad de estas: relaciones topológicas, relaciones de distancia, relaciones de orientación, etc. A esto ha de su-

marse la diferente naturaleza de las relaciones espaciales en el lenguaje habitual, que es eminentemente cualitativa, a diferencia de los procesos que se implementan en un SIG, de naturaleza cuantitativa.

Modelizar estas relaciones de forma correcta e integrar todos los puntos de vista es importante para hacer de los SIG herramientas de análisis completas en las que puedan expresarse de forma intuitiva y coherente todas las relaciones existentes.

Consultas y operaciones con bases de datos

12.1. Introducción

El análisis más simple que podemos efectuar sobre una capa (o varias) de información geográfica es la simple consulta de esta. Entendemos por consulta una operación en la cual *preguntamos* a los datos geográficos algún tipo de cuestión simple, generalmente basada en conceptos formales sencillos. Este tipo de análisis, aunque no implica el uso de conceptos analíticos complejos, es uno de los elementos clave de los SIG, pues es parte básica del empleo diario de estos.

En el contexto espacial, una consulta representa un uso similar al que damos a un mapa clásico, cuando en base a este respondemos a preguntas como *¿qué hay en la localización X?* o *¿qué ríos pasan por la provincia Y?* No obstante, no debemos olvidar que los datos espaciales tienen dos componentes: una espacial y otra temática. Preguntas como las anteriores hacen referencia a la componente espacial, pero igualmente pueden efectuarse consultas que se apliquen sobre la parte temática. Y más aún, pueden efectuarse consultas conjuntas que interroguen a los datos geográficos acerca de los atributos espaciales y temáticos que estos contienen.

Las consultas se entienden en general como relativas a capas vectoriales, pues son dicho modelo de representación y su estructura de datos los que mejor se adaptan a la forma particular de las consultas. En este capítulo veremos cómo trabajar mayoritariamente con datos vectoriales, aunque también se harán referencias a datos ráster, ya que estos últimos contienen igualmente datos geográficos y pueden consultarse y responder a preguntas como las formuladas anteriormente.

En las capas vectoriales, y como vimos en los capítulos 5 y 8, la división entre la componente temática y espacial es más patente, existiendo incluso una división a nivel de archivos y de los elementos tecnológicos empleados para el trabajo con cada una de ellas dentro de un SIG. En el caso de las consultas, se mantiene un enfoque similar, y encontramos esa misma separación. Los lenguajes de consulta, que en breve veremos con más detalle y que resultan básicos para elaborar consultas y obtener resultados, han seguido una evolución paralela a la de los propios sistemas gestores de bases de datos en su adaptación al entorno espacial de los SIG.

Siguiendo este mismo enfoque, estudiaremos en primer lugar los conceptos fundamentales relativos a consultas en bases de datos, sin tratar por el momento la componente espacial. Posteriormente extenderemos estos conceptos para ver la verdadera potencia de estas dentro del ámbito SIG, que resulta de añadir la componente espacial y los conceptos sobre relaciones espaciales que vimos en el capítulo 11.

Si estás familiarizado con los conceptos relativos a bases de datos no espaciales y las consultas sobre estas, puedes prescindir de leer la próxima

sección y avanzar hasta la siguiente para ver directamente las particularidades del trabajo con bases de datos espaciales. De cualquier modo, el capítulo no pretende ser un manual sobre el uso de bases de datos o sus fundamentos, ya que es un tema muy amplio y escapa por completo al alcance de este texto.

12.2. Consultas en un SIG

Si partimos de datos vectoriales y de la presencia de algún sistema gestor de bases de datos o tecnología similar dentro de un SIG, una consulta no es sino una llamada a dicho sistema gestor, el cual devuelve como respuesta una serie de elementos tomados de la información contenida en la base de datos. Es decir, del total de datos obtenemos como consecuencia de la consulta una parte de los mismos. La respuesta a nuestra consulta es un conjunto de elementos, de la misma forma que si en un mapa impreso preguntamos *¿qué hay aquí?* y obtenemos como respuesta los datos correspondientes al punto que señalamos. Estos datos son una fracción particular del conjunto de todos los contenidos en dicho mapa.

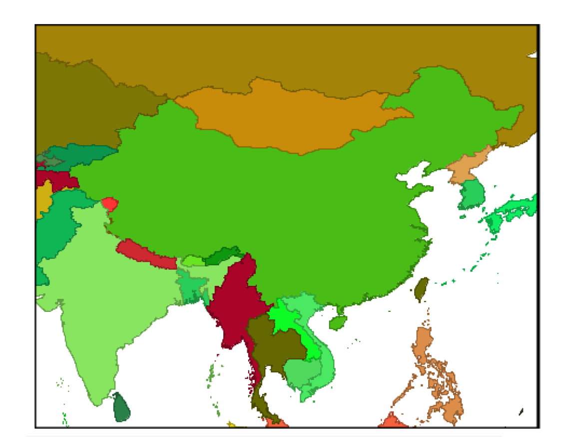
El resultado de una consulta en un SIG generalmente es lo que conocemos como *selección*. De todos los registros de la tabla de datos, aquellos que cumplen el criterio indicado se marcan como seleccionados, y posteriormente pueden utilizarse únicamente estos como base de otro análisis, o simplemente el usuario puede ver cuáles han sido los seleccionados para así obtener la respuesta a su consulta.

Como veremos más en detalle en las siguientes secciones, las consultas pueden hacerse sobre la componente temática de los datos, sobre la espacial, o sobre ambas. En cualquier caso, sabemos ya que estas en un SIG se encuentran vinculadas, con lo que el resultado de la consulta afecta a ambas. La selección se hace patente sobre ambas componentes, con independencia de cuál de ellas haya sido la encargada de aplicar el criterio de selección. En el entorno habitual de un SIG, con su interfaz gráfica, tanto la tabla de atributos como la representación visual de la componente espacial se ven afectadas por la realización de una consulta. La figura 12.1 muestra gráficamente este hecho.

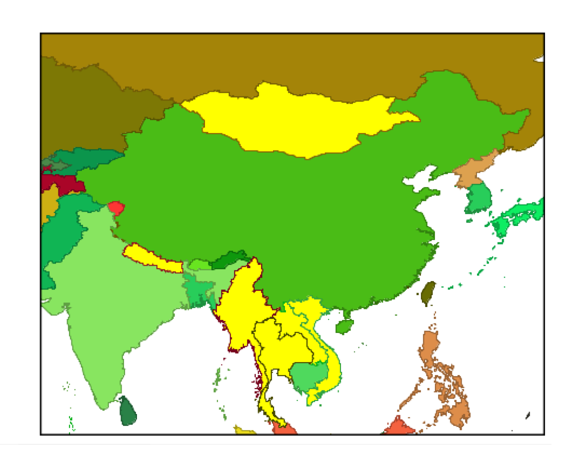
Esta presencia gráfica es importante dentro del entorno de los SIG, tanto para mostrar el resultado de las consultas como para ayudar en la formulación de estas. En contraste con el carácter textual de una base de datos, el empleo de dichas bases de datos y la realización de consultas en un SIG incorpora una representación gráfica que resulta esencial [38]

Junto con la selección de entidades dentro de una capa existente, una consulta nos vale también para extraer información de una base de datos de acuerdo a nuestras necesidades, y para crear posteriormente y a partir de dicha información una nueva capa. Esta operación es útil cuando la base de datos de la que disponemos es muy voluminosa y solo resulta de interés para nuestro trabajo una parte de ella. Puede tratarse de una parte en el sentido espacial (la base de datos contiene datos a nivel mundial y se quiere trabajar a nivel estatal), en el sentido temático (la base de datos contiene mucha información de cada entidad y solo interesan algunos campos), o en una combinación de ambas. Para extraer dicha parte y trabajar únicamente con ella, utilizaremos una consulta.

Así, la selección de una serie de entidades dentro de una capa o la extracción de dichas entidades de la base de datos para la creación de dicha capa son dos aplicaciones habituales de las consultas que seguidamente veremos.



Pais	Poblacion	Área
Birmania	43,09	669820
China	1275,2	9337880
India	894,6	3089282
Laos	47,22	230556
Mongolia	2,22	1559176
Nepal	19,92	147292
Tailandia	57,32	515144
Vietnam	71,21	327122



Pais	Poblacion	Área
Birmania	43,09	669820
China	1275,2	9337880
India	894,6	3089282
Laos	47,22	230556
Mongolia	2,22	1559176
Nepal	19,92	147292
Tailandia	57,32	515144
Vietnam	71,21	327122

Figura 12.1: El resultado de una consulta temática en un SIG es una selección de entidades, que implica tanto a la componente temática como a la espacial de cada una de ellas. En ambos casos, el color amarillo indica los elementos seleccionados.

12.3. Consultas temáticas

Comencemos analizando algunas consultas temáticas y viendo cómo, aunque se realicen en base a datos espaciales como los que utilizamos en un SIG, en realidad en ellas la componente espacial no se emplea. Sea por ejemplo una capa con los distintos países del mundo y una serie de valores económicos y sociales asociados a cada uno de ellos. Consideremos las siguientes preguntas:

- ¿Qué países tienen un Producto Interior Bruto mayor que el de España?
- ¿Qué países han experimentado un crecimiento económico en el último año?
- ¿Cuántos países tienen más de 200 millones de habitantes?

En todos estos casos estamos haciendo referencia a países, los cuales, como sabemos, estarán asociados a elementos geométricos que definan sus propiedades espaciales, es decir, a una componente espacial. Esta componente es la que permite que, además de poder plantear las consultas anteriores, podamos representar cada país en la pantalla y visualizarlo, o saber cuáles de ellos se encuentran en el hemisferio norte (esta sería una consulta espacial, de las que más adelante en este mismo capítulo veremos).

Sin embargo, cuando realizamos consultas como las tres anteriores, no acudimos a la componente espacial. Consultas como estas podrían resolverse si en lugar de una capa dentro de un SIG tuviéramos, por ejemplo, un simple anuario estadístico lleno de tablas con datos correspon-

dientes a cada país. De hecho, antes del desarrollo de los SIG, ese tipo de datos, aunque referidos a elementos geográficos, se almacenaban en documentos tales como dicho anuario, y no específicamente en mapas. Es fácil encontrar mapas del mundo con meras divisiones fronterizas entre países (un mapa político) o quizás con elevaciones y elementos orográficos (un mapa físico), pero no es tan sencillo adquirir un mapa en el que pueda conocerse el crecimiento económico del ultimo año en cada país. Esta información se puede adquirir, sin embargo, de forma sencilla en ese anuario estadístico que citamos.

Antes de la aparición de los SIG, la componente temática (el anuario estadístico) y la espacial (el mapa político) iban por separado. Hoy en día, y gracias a los SIG, podemos trabajar con ellas de forma conjunta, pues es fácil ver que existe una relación entre ambas.

Por un momento, dejemos de lado la componente espacial de cada país, y pensemos que solo conocemos de él algunas variables socio–económicas tales como el PIB, la población, el idioma que se habla o el nombre de su moneda, tal y como se recogen en la tabla de la figura 12.1

12.3.1. Mecanismos de consulta y operaciones básicas

Consultas como las anteriores pueden expresarse fácilmente en un idioma tal como el español y son de igual modo fácilmente entendibles por cualquiera que conozca este idioma. El problema es que el ordenador, y por tanto el Sistema de Información Geográfica, no entiende estas expresiones, siendo necesario formular las consultas de alguna forma que pueda ser interpretada correctamente por el SIG o el gestor de bases de datos correspondiente.

Dentro de un SIG hay muchas formas de expresar una consulta. Una forma simple es a través de expresiones lógicas relativas a los campos de la tabla de atributos. Planteando las consultas como expresiones condicionales, la respuesta a estas son aquellas entidades que hacen verdadera dicha expresión.

Para trabajar desde este punto en adelante, vamos a suponer que disponemos de una tabla con datos de países del mundo, la cual contiene los siguientes campos:

- `NOMBRE`
- `CAPITAL`
- `MONEDA`
- `POBLACION_ACTUAL`
- `POBLACION_ANTERIOR`
- `SUPERFICIE`

Por ejemplo, para saber el número de países con población mayor de 200 millones, podríamos utilizar una expresión como la siguiente: `'POBLACION_ACTUAL'` $>$ `200000000`. Para saber en qué países aumentó la población en el ultimo año, y puesto que disponemos además de un campo con la población de año anterior, podemos plantear una expresión de la forma `POBLACION_ACTUAL >POBLACION_ANTERIOR`.

Estas expresiones condicionales se conocen con el nombre de *predicados*.

Los predicados no han de ser necesariamente de carácter numérico. Por ejemplo, para saber qué países pertenecen a la Unión Económica Europea podríamos hacerlo mediante el predicado `MONEDA = 'Euro'`.

Consultas de esta índole también pueden efectuarse si los datos geográficos se basan en un modelo de datos ráster. En este caso, podemos de igual modo ver qué celdas cumplen una condición dada como, por ejemplo, tener un valor mayor que un valor predefinido. Sin embargo, este tipo de operaciones no se suelen ver habitualmente como consultas, sino como operaciones de lo que se conoce como *álgebra de mapas*, en particular una operación denominada *reclasificación*. Vere-

mos con detalle estas operaciones en el capitulo 15, enteramente dedicado a ellas.

Las consultas mediante expresiones condicionales pueden ser algo más complejas que lo que hemos visto hasta ahora, ya que pueden hacer referencia a varios campos. Por ejemplo, para responder a cuestiones como las siguientes:

- ¿Qué países de la zona euro tienen más de 40 millones de habitantes?
- ¿En qué países de habla inglesa aumentó la población durante el último año?

Para expresar esas consultas se han de incluir elementos de la denominada *lógica booleana*[1]. Para entender de forma sencilla este concepto, podemos reescribir las consultas anteriores de la siguiente manera:

- ¿Qué países tienen como moneda el euro *y* tienen más de 40 millones de habitantes?
- ¿Que países hablan inglés *y* experimentaron un aumento de población durante el último año?

La partícula *y* nos indica que realmente nuestra consulta se compone de dos condiciones. Por ejemplo, en el primero de los casos se debe cumplir la condición `Moneda = 'Euro'`, y al mismo tiempo la condición `POBLACION_ACTUAL > 40000000.` La sintaxis habitual para expresar esto a la hora de formular la consulta es emplear el termino inglés `AND`, de forma que tendríamos la expresión `MONEDA = 'Euro' AND POBLACION_ACTUAL > 40000000`.

Otros operadores lógicos que podemos emplear son el operador disyuntivo *o* (`OR`) o el operador de negación (`NOT`). Añadiéndolos a las expresiones condicionales podemos crear consultas más elaboradas para extraer la información que buscamos.

En realidad, formular una consulta de este tipo es buscar un subconjunto particular (el formado por las entidades que quedarán seleccionadas) dentro de un conjunto global (la capa geográfica). Por ello, es útil plantear estas operaciones lógicas desde el punto de vista de la teoría de conjuntos, y hacer uso de los denominados *diagramas de Venn*[2], que muestran de forma gráfica y muy intuitiva el significado de dichas operaciones. En la figura 12.2 pueden verse los diagramas correspondientes a las operaciones que hemos visto hasta el momento.

Más adelante volveremos a encontrar esquemas similares a estos, en particular al estudiar las operaciones de solape entre capas vectoriales, en el capítulo 19.

Una operación muy habitual en el ámbito de las bases de datos es la *unión* de tablas. Si la componente temática en el modelo vectorial se almacena en una tabla de atributos, es posible, mediante esta operación, almacenar dicha componente en un conjunto de ellas interrelacionadas, lo cual en términos generales conlleva una mejor estructuración, como ya vimos en el capítulo dedicado a las bases de datos. En otras palabras, resulta conveniente no poner toda la información en una tabla, sino dividirla en un conjunto adecuadamente estructurado de tablas, que después pueden combinarse y utilizarse de la manera que resulte más apropiada.

Lógicamente, si la información se encuentra disponible en una serie de tablas, pero las entidades geométricas que contienen la componente espacial del dato solo tienen asociada una de ellas, es necesario algún elemento adicional que permita relacionar a todas esas tablas. Este elemento son las operaciones de unión, que pueden efectuarse entre varias capas (cada una aporta su tabla de atributos) o también con tablas aisladas, sin necesidad de que estas tengan asociada una componente espacial.

[1]Denominada así por el matemático irlandés George Boole(1815, 1864)
[2]John Venn (1834, 1923), un matemático inglés, fue el primero en proponer este tipo de diagramas en 1880

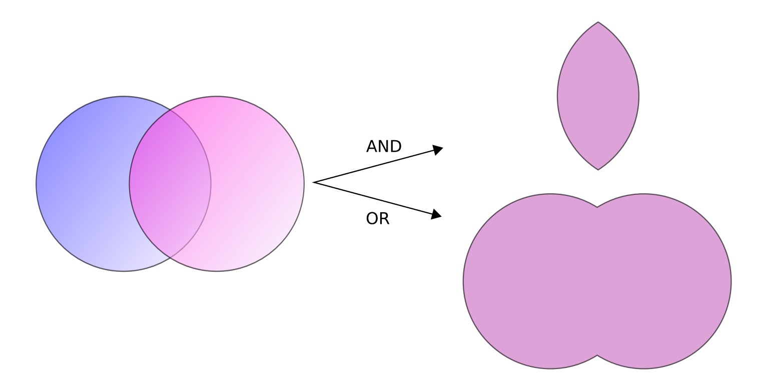


Figura 12.2: Diagramas de Venn.

Veamos un ejemplo de una unión entre tablas para comprender el significado de esta operación. En el cuadro 12.1 se muestran dos tablas. En la primera de ellas, que supondremos que se encuentra asociada a una capa de polígonos, encontramos un valor numérico que identifica cada entidad (`ID`) y es distinto para cada una de ellas, y otro valor numérico también entero (`TIPO_SUELO`) que nos indica el grupo de tipo de suelo de dicho polígono. Es decir, es la tabla de una capa de usos de suelo. En la segunda tabla, que es una tabla no asociada a ninguna capa, encontramos la equivalencia entre los valores de tipo de suelo y el nombre de dicho tipo de suelo (`NOMBRE_SUELO`), así como en el campo denominado `APTITUD`, que, en una escala de 1 a 10, clasifica la aptitud del suelo para el crecimiento de una determinada especie vegetal.

En estas tablas existe un campo común, que es el que contiene el código numérico del tipo de suelo (en el ejemplo tiene el mismo nombre en ambas tablas, pero esto no ha de ser necesariamente así, basta con que la variable que contengan ambos campos sea la misma), y que podemos emplear para establecer la relación entre las dos tablas. Si un polígono tiene asociado un suelo de tipo 1, y gracias a la tabla adicional sabemos que el suelo de tipo 1 es un fluvisol y que su aptitud es 5, entonces podemos decir que dentro de ese polígono el suelo es de tipo fluvisol y tiene dicha aptitud, aunque en la tabla asociada no se encuentre directamente esta información. Esta forma de proceder debe resultar ya familiar, pues la unión se basa en el uso de un atributo clave, que ya vimos en el capítulo 8 dedicado a las bases de datos.

Al efectuar una unión, *pasamos* la información de una tabla a la otra, en la medida en que esto sea coherente con las coincidencias existentes entre ellas. El resultado es una nueva tabla que extiende la tabla original, incorporando información de otra tabla adicional. En el cuadro 12.2 puede verse la tabla resultante de la anterior unión[3].

Mantener la información dividida en varias tablas de la forma anterior tiene muchas ventajas, como ya en su momento comenzamos a ver en el apartado 8.2.6 cuando tratamos el diseño de

[3]Esta tabla es la resultante de uno de los tipos de unión posibles, la denominada unión *natural*. Otros tipos de unión para las mismas tablas de partida dan como resultado tablas distintas, aunque no se usan con tanta frecuencia. La unión *cartesiana* genera una tabla que contiene todas las combinaciones posibles entre elementos de las dos tablas implicadas. En otros casos, la tabla generada puede contener los dos campos utilizados como enlace, uno por cada tabla de las utilizadas en la unión

ID	TIPO_SUELO
1	3
2	1
3	3
4	3
5	2

TIPO_SUELO	NOMBRE_SUELO	APTITUD
1	Fluvisol	5
2	Cambisol	7
3	Leptosol	4

Cuadro 12.1: Unas sencillas tablas para efectuar una unión en función de su campo común `TIPO_SUELO`.

ID	TIPO_SUELO	NOMBRE_SUELO	APTITUD
1	3	Leptosol	4
2	1	Fluvisol	5
3	3	Leptosol	4
4	3	Leptosol	4
5	2	Cambisol	7

Cuadro 12.2: Resultado de la unión de tablas en base a su campo común `TIPO_SUELO`.

bases de datos, y ahora al estudiar las consultas vemos plasmadas de modo claro una buena parte de ellas. El hecho de codificar cada nombre de tipo de suelo con un valor numérico hace más sencillo el introducir este parámetro, pues evita teclear todo el nombre cada vez que se quiera añadir algún nuevo registro a la tabla principal (en este caso, la que se encuentra asociada a la capa). Esto además disminuye la probabilidad de que se cometan errores al introducir dichos valores y garantiza la homogeneidad de nombre, pues estos solo se encuentran almacenados una única vez en la tabla adicional. Las restantes tablas se *alimentan* de esta.

Las ventajas son mayores si se piensa que la tabla que caracteriza cada tipo de suelo puede estar unida no a una sola tabla de atributos, sino a un número mayor de ellas. En tal caso, ampliar o modificar la información de las tablas resulta mucho más sencillo, pues solo se deben efectuar dicha ampliación o modificación sobre la tabla de la que las demás toman sus datos. Si por ejemplo quisiéramos modificar la aptitud del tipo de suelo *leptosol* de 4 a 5, basta realizar esta modificación en la tabla auxiliar. El número de veces que este tipo de suelo aparece en la tabla de atributos de la capa resulta indiferente a la hora de efectuar esta modificación, pues solo ha de cambiarse una vez.

Si se desean añadir nuevos campos tales como el nombre del tipo de suelo en un idioma distinto o la aptitud de cada tipo de suelo para una especie distinta, estos no han de añadirse a cada tabla de atributos, sino solo a la tabla auxiliar. Por otra parte, el almacenamiento estructurado tiene como resultado una información menos redundante, y por tanto un menor volumen de los datos. En definitiva, existen muchas ventajas asociadas a una estructuración adecuada de los datos, las cuales pueden aprovecharse definiendo las relaciones entre todas esas tablas a través de operaciones tales como la unión.

Todas las operaciones que hemos visto se realizan de forma diferente según el SIG que empleemos, pues constituyen herramientas independientes que se implementan de una u otra forma dependiendo del producto. Existe, no obstante, una forma unificada de llamar a estas y a otras funciones, y es a través de los *lenguajes de consulta*. Los lenguajes de consulta son un elemento fundamental de las bases de datos y, básicamente, y como su nombre indica, se trata de lenguajes pensados para poder expresar todo tipo de consultas relativas a una base de datos y obtener así una información dada a partir de ella. Es decir, permiten expresar todas las consultas que hasta el momento hemos visto en este capítulo, así como otras más complejas y elaboradas.

En realidad, son lenguajes que buscan dar solución a todas las necesidades de trabajo con bases de datos, y estas incluyen no solo aquellas relacionadas con consultas (aunque representen la operación más habitual) sino también las que derivan del mantenimiento y creación de dicha base de datos. En su empleo más habitual, los lenguajes de consulta han de ofrecer una forma sencilla y eficaz de que un usuario cualquiera pueda efectuar consultas sobre una base de datos, formulando estas de una forma lógica y flexible.

Un lenguaje de consulta posee una sintaxis rígida, que lo asemeja a un lenguaje de programación (de hecho, se trata de un lenguaje de programación como tal). No obstante, la complejidad algorítmica inherente a la propia consulta queda oculta en la expresión, asemejándose en ese sentido más a un lenguaje natural. Un lenguaje de consulta no sirve para implementar algoritmos, sino que expresa de una forma más natural (aunque con una sintaxis adecuada al entorno computacional en que se encuentra) dichos algoritmos de consulta.

Dicho de otro modo, estos lenguajes de consulta van a expresar en líneas generales *qué* es lo que se quiere hacer, pero no *cómo* se debe hacer, al contrario que los lenguajes de programación (tales como los que se emplean para programar, por ejemplo, un SIG), que permiten describir formalmente algoritmos y procedimientos[4].

El siguiente paso es, pues, estudiar cómo los lenguajes de consulta se incorporan a un SIG y la forma de utilizar estos. El lenguaje de consulta más extendido para bases de datos relacionales, tanto dentro como fuera del ámbito SIG, es el denominado SQL (acrónimo de *Structured Query Language* o *Lenguaje de Consulta Estructurado*).

12.3.2. El lenguaje SQL

El lenguaje SQL es un lenguaje de consulta pensado para el manejo de datos, e incluye elementos para realizar todas aquellas operaciones habituales que se presentan en el uso de una base de datos. Su utilización es habitual dentro de cualquier sistema que implique el manejo de datos mediante un gestor de bases de datos, y un cierto conocimiento de sus fundamentos es de indudable interés para el usuario de SIG. El objetivo de esta parte no es constituir una referencia de este lenguaje, sino tan solo presentar sus principales elementos para mostrar la forma en que un lenguaje de consulta soluciona las necesidades que operaciones como las vistas hasta este punto plantean dentro de un SIG. Las referencias sobre SQL son muy abundantes y el lector interesado no tendrá dificultad en encontrar docenas de libros sobre este tema. Una referencia completa es [105].

Podemos distinguir tres componente principales dentro del SQL:

[4]Los lenguajes de programación se dice que son lenguajes *procedurales*, mientras que los lenguajes de consulta se denominan *no procedurales*

- Un lenguaje de definición de datos. Mediante él podemos definir las características fundamentales de los datos y sus relaciones.
- Un lenguaje de manipulación de datos. Permite añadir o modificar registros a las tablas de la base de datos. Las funciones de consulta también se incluyen dentro de este lenguaje
- Un lenguaje de control de datos. Sus comandos permiten controlar aspectos como el acceso a los datos por parte de los distintos usuarios, así como otras tareas administrativas

En principio, es la segunda componente —el lenguaje de manipulación de datos— la que resulta aquí de mayor interés, y a la que acudiremos de forma más frecuente, pues contiene los elementos de consulta.

La forma en que se realizan consultas a una base de datos empleando el lenguaje SQL es a través de *sentencias* en dicho lenguaje. Una sentencia SQL de consulta tendrá habitualmente una forma como la siguiente:

```
SELECT lista_de_columnas
FROM nombre_de_tabla
WHERE expresión_condicional
```

`Lista_de_columnas` es una lista con los nombres de los campos que nos interesa incluir en la tabla resultante de la consulta, `nombre_de_tabla` es el nombre que identifica la tabla de la que queremos tomar los datos, y `expresión_condicional` un predicado que establece la condición que han de cumplir los registros a tomar.

Como vemos, no es muy distinto de lo que hemos visto hasta ahora, y simplemente es otra forma de plantear consultas, de modo similar a lo que conocemos. La flexibilidad del lenguaje y la adición de funciones y nuevas ordenes permiten, no obstante, expresar prácticamente cualquier consulta, por compleja que esta sea, y extraer todo el potencial de la base de datos.

Sin animo de resultar un repaso exhaustivo a todos los elementos del lenguaje (ello requeriría mucha más extensión que todo este capítulo), a continuación se muestran algunos ejemplos de expresiones SQL explicados después en lenguaje natural, para dar una idea de la forma y capacidades del lenguaje. Una vez más, se anima al lector interesado a consultar las referencias propuestas para mayores detalles. Para estos ejemplos utilizaremos la tabla de países cuya estructura ya hemos presentado al comienzo de este capítulo.

```
SELECT *
FROM Paises
WHERE Moneda = 'Euro'
AND Poblacion_actual > 40000000
```

Esta consulta recupera todos aquellos registros en los que la población actual supera los 40 millones y la moneda es el euro. El asterisco indica que, para cada uno de estos registros, deben recuperarse todos los campos existentes. Se podría formar una nueva tabla solo con los nombres de los países que cumplen la condición establecida, mediante la siguiente sentencia:

```
SELECT Nombre
FROM Paises
WHERE Moneda = 'Euro'
AND Poblacion_actual > 40000000
```

Las consultas pueden ser mucho más complejas, y las sentencias `SELECT` pueden usarse como parámetros dentro de un predicado lógico. Por ejemplo, supongamos que disponemos de una tabla denominada `Capitales` con datos de todas las capitales del mundo, y que cuenta con los siguientes campos:

```
- Nombre
- Poblacion_hombres
- Poblacion_mujeres
```

La sentencia siguiente recupera en una nueva tabla todos los países cuyas capitales tienen una población de hombres mayor que de mujeres.

```
SELECT Nombre FROM Paises
WHERE Capital IN
   (SELECT Nombre
    FROM Capitales
    WHERE Poblacion_hombres
       > Poblacion_mujeres)
```

La subconsulta entre paréntesis crea una tabla con los nombres de las capitales que cumplen la condición relativa a las poblaciones de hombres y mujeres. La otra consulta selecciona los países cuya capital aparece en dicha tabla (esa inclusión la define el comando `IN`). En conjunto, tenemos una única consulta, pero que se basa en dos tablas con una relación entre sí.

Los campos `Capital` en la tabla `Países` y `Nombre` en la tabla `Capitales` son los que establecen la relación entre ambas tablas, permitiendo unir la información de estas. No obstante, el resultado de la expresión anterior no es una unión tal y como la hemos visto. Para realizar una unión mediante SQL podemos utilizar una expresión como la siguiente:

```
SELECT *
FROM Paises, Capitales
WHERE Paises.Capital = Capitales.Nombre
```

Como se dijo, las uniones no se limitan un tipo particular de unión como el que vimos. SQL incluye el comando `JOIN`, el cual permite trabajar con todo ese abanico de distintas uniones.

Además de lo anterior, pueden emplearse operadores para que la tabla que constituye la respuesta a la consulta contenga campos adicionales calculados en función de los existentes en la tabla origen. Por ejemplo:

```
SELECT Nombre, Poblacion / Area
AS Densidad
FROM Paises
WHERE Moneda = 'Euro'
```

Esta consulta recupera todos los países donde la moneda utilizada es el Euro, y para cada uno de ellos define dos atributos: el nombre (directamente obtenido de la propia tabla de países) y la densidad (en un campo denominado `DENSIDAD`, calculado como el cociente entre la población y el área).

En resumen, el lenguaje SQL permite expresar todo tipo de consultas y hacerlo de forma sistemática y relativamente sencilla, de una forma bastante similar a como lo haríamos en un lenguaje natural.

12.4. Consultas espaciales

Ahora que ya sabemos cómo sacar partido de los atributos (es decir, la componente temática), es hora de incorporar la componente espacial que se asocia a estos. A las consultas que pusimos como ejemplo en la sección anterior, podemos añadir otras como las siguientes:

- ¿Qué países comparten frontera con Alemania?
- ¿Cuántos países se encuentran completamente en el hemisferio sur?
- ¿Qué países están a menos de 2000 km de España?

Para dar respuesta a esas cuestiones, basta analizar la componente espacial sin necesidad de recurrir a los datos temáticos con los que hemos trabajado anteriormente. Son consultas puramente espaciales. Aunque estas consultas amplían lo que ya conocemos, en realidad no abren ninguna nueva vía de estudio de los datos geográficos. Son consultas a las que podríamos responder uti-

lizando un mero mapa impreso, sin aprovechar el hecho de que, como hemos visto, dentro de un SIG las componentes espacial y temática se hallan íntimamente vinculadas.

La verdadera potencia de las consultas espaciales la encontramos en la combinación de estas consultas sobre la componente espacial y las que vimos anteriormente sobre la componente temática. Así, se pueden plantear, por ejemplo, cuestiones como:

- ¿Qué países del hemisferio norte tiene una densidad de población mayor que la de Perú?
- ¿Cuántos países con más de 10 millones de habitantes se encuentran a menos de 1000 km de la frontera de Rusia?

Estas consultas incorporan elementos que hacen necesario acudir a la tabla de atributos, y otros que requieren analizar la componente espacial, estudiando las relaciones espaciales y topológicas de las geometrías asociadas.

Los lenguajes de consulta pensados para el trabajo exclusivo con datos no espaciales no permiten formular consultas que incorporen elementos espaciales, y por lo tanto no resultan suficientes para expresar las anteriores cuestiones. Tanto las bases de datos como los lenguajes de consulta son válidos para analizar la componente temática, pero no para el análisis global de ambas componentes tal y como este ha de llevarse a cabo dentro de un SIG, por lo que es necesario añadir elementos adicionales.

Sin embargo, no es solo mediante un lenguaje de consulta como podemos plantear dichas consultas espaciales a través de un SIG. Al igual que en el caso de la componente temática, a la hora de efectuar consultas sobre la componente espacial o bien sobre ambas conjuntamente, existen diversas formas de plantear dichas consultas, algunas de ellas mucho más inmediatas y sencillas. En el caso particular de la componente espacial, y por la propia naturaleza de esta, que puede ser representada gráficamente, la forma más simple de efectuar una consulta es, precisamente, de forma gráfica.

Este es el mismo mecanismo que emplearíamos a la hora de trabajar con un mapa impreso clásico. Si señalamos sobre nuestro mapamundi y preguntamos ¿qué país es *este*?, estamos estableciendo física y visualmente el criterio de consulta con nuestro propio dedo. Dentro de un SIG, podemos hacer clic con el ratón (nuestro dedo dentro de dicho SIG) en un determinado punto de la representación en pantalla de una capa geográfica, y realmente estamos diciendo: ¿qué entidad de la capa es la que hay *aquí*? o ¿qué entidad es *esta*?

Al hacer esto, estamos empleando las relaciones espaciales que veíamos en el capítulo 11, y en particular en este caso la inclusión de un punto dentro de un polígono. Al efectuar la consulta, el SIG comprueba si el punto definido por nuestro clic de ratón se encuentra dentro de los polígonos que representan cada país. Si eso es así, el país en cuestión queda seleccionado.

Una vez más, no debe pensarse que esta consulta puntual es exclusiva de los datos vectoriales. Podemos igualmente ir a una localización dada y preguntar por lo que hay en dicha localización con independencia del modelo de datos. Una capa ráster nos devolverá sencillamente el valor en la celda que cae en el emplazamiento señalado. Si la capa posee varias bandas, tal como una imagen multiespectral, nos devolverá un vector de valores correspondientes a los valores de todas las bandas en dicho punto.

Como veremos en el capítulo 24 dedicado a servicios remotos, algunos de estos servicios nos permiten realizar consultas igualmente sobre datos ráster y coberturas. En cada punto de la cobertura tenemos una información compleja, que podemos recuperar del mismo modo que para otro tipo de capas, sin más que preguntar a dicha

cobertura acerca los datos correspondientes a un punto dado.

La consulta sobre capas no vectoriales es, sin embargo, menos interesante, pues el mayor interés aparece cuando consideramos entidades en el modelo geográfico y efectuamos consultas sobre las propiedades espaciales de dichas entidades. El modelo vectorial es el mejor adaptado a las consultas, no solo cuando trabajamos con la componente temática, como ya vimos, sino igualmente cuando se trata de consultas puramente espaciales.

La consulta sobre un punto concreto que hemos descrito la incorporan la gran mayoría de los SIG y es una herramienta de primer orden, sumamente sencilla, que nos permite hacer un uso simple aunque muy práctico de los datos geográficos. No obstante, una consulta espacial de este tipo puede ser más compleja e incorporar en el criterio algo más que un único punto. Por ejemplo, podemos seleccionar todas las entidades dentro de un área rectangular, o bien dentro de un polígono cualquiera que podríamos definir directamente sobre la propia representación en pantalla (Figura 12.3).

Las relaciones que utilizamos en este caso ya no son entre punto y polígono, sino entre polígonos. La selección puede incluir tanto los países que se encuentran por completo contenidos dentro del polígono, como aquellos que intersecan con este. Es decir, que podemos aplicar varias de las relaciones que en su momento estudiamos entre dos polígonos.

Además de poder efectuar estas consultas con un elemento tal como un punto o un polígono, también podemos valernos de otra capa con sus propias geometrías. Por ejemplo, si disponemos del contorno del continente europeo, podemos consultar la capa de países del mundo y ver cuáles se encuentran en Europa. O una capa de ríos del mundo nos valdría para responder a la pregunta «¿qué países atraviesa el Nilo?»

Las relaciones espaciales entre las entidades de varias capas pueden emplearse para efectuar una selección, pero también para otra de las operaciones importantes que veíamos en el caso de la componente temática: la unión. En aquel caso, se establecían las relaciones entre tablas de acuerdo a un predicado relativo a la propia información de las tablas (en la mayoría de los casos, que los valores de dos campos, uno en cada tabla, fueran coincidentes). La incorporación de la componente espacial implica la aparición de predicados espaciales, y estos pueden emplearse del mismo modo que los no espaciales para definir un criterio de unión.

Un ejemplo muy sencillo de unión espacial es el que encontramos si combinamos la capa de países del mundo que venimos utilizando con una capa de ciudades del mundo. Podemos unir a la tabla de esta segunda capa todos los valores que caracterizan al país al que pertenece cada ciudad. Si existe un campo común entre ambas tablas de atributos (por ejemplo, el nombre del país), esto serviría para efectuar esta unión. No obstante, esto no es necesario, ya que existe otro elemento común que no se encuentra almacenado dentro de la tabla, pero que puede tomarse de la componente espacial: toda ciudad debe estar situada dentro de los límites del país al que pertenece. Esto sirve para establecer la relación entre las tablas, y cada ciudad debe relacionarse con aquella entidad dentro de cuya geometría se encuentre el punto que la representa.

De modo similar a como ocurría con las operaciones temáticas, todas estas operaciones pueden llevarse a cabo en un SIG mediante herramientas sencillas que se encargan de efectuar las selecciones o uniones, utilizando tanto elementos gráficos como textuales. Disponemos así de herramientas de consulta temática y herramien-

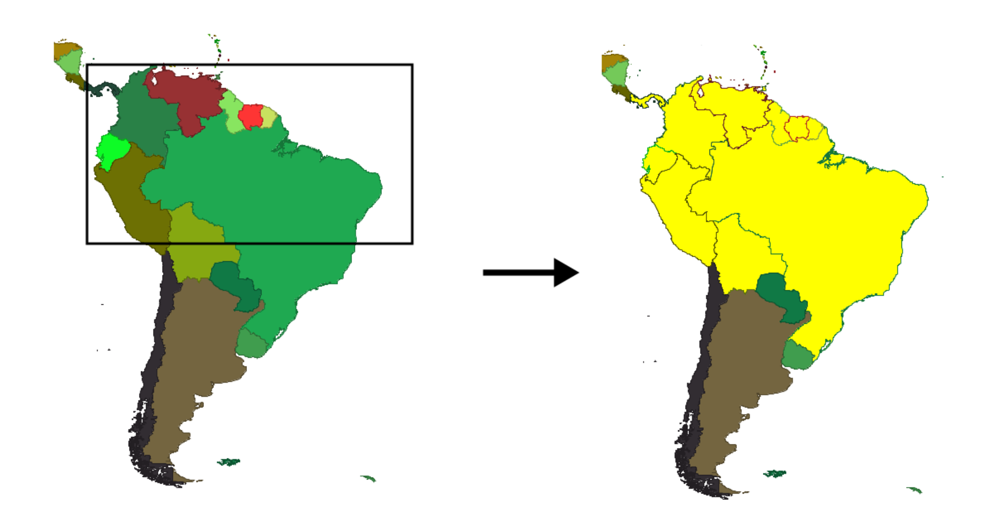


Figura 12.3: Consulta mediante rectángulo. Los países que intersecan con los limites definidos por dicho rectángulo quedan seleccionados.

tas de consulta espacial, ambas como utilidades independientes. Podemos, sin embargo, dotar de mayor potencia a las realización de consultas si combinamos ambas componentes del dato geográfico.

Es en este punto donde los lenguajes de consulta que ya hemos visto hacen su aparición. Estos lenguajes han demostrado ser idóneos para el manejo de las bases de datos, y resulta lógico pensar en ellos como base para un lenguaje más potente que permita incorporar la componente espacial a las consultas. Tenemos de este modo los *lenguajes de consulta espacial*

12.4.1. Lenguajes de consulta espacial

Los lenguajes de consulta espacial son la extensión lógica de los lenguajes de consulta con objeto de adaptarse al manejo de datos espaciales. Del mismo modo que las bases de datos han de adaptarse para almacenar los datos espaciales, como ya vimos en el capítulo dedicado a estas, los lenguajes de consulta deben hacer lo propio para ser capaces de recoger aquellas consultas que hagan un uso explícito de las propiedades espaciales de los objetos almacenados.

El lenguaje SQL que ya conocemos no resulta suficiente para expresar algunas de las consultas presentadas en el apartado anterior, pero sin embargo sigue siendo de utilidad para consultas no espaciales. Las extensiones al lenguaje SQL constituyen la forma más inmediata de obtener un lenguaje de consulta espacial adecuado para un uso completo, pues combinarán nuevas capacidades de consulta espacial con aquellas de tipo no espacial del lenguaje SQL, probadamente robustas.

Un primera solución aparece con la revisión del lenguaje SQL llevada a cabo en 1999 (conocida como SQL:1999 o SQL3), en la cual se permite la creación de tipos personalizados de datos. Frente a los tipos originales tales como enteros, cadenas o valores booleanos, todos ellos poco adecuados para almacenar objetos espaciales, existe ahora la posibilidad de crear tipos más acordes con la naturaleza espacial de la información almacena-

da. SQL3 está orientado a objetos, y cada tipo puede tener una serie de métodos asociados, lo cual facilita la realización de consultas complejas.

El problema radica en que la propia flexibilidad de este mecanismo no favorece la unicidad necesaria para la interoperabilidad de los datos. Un mismo tipo puede implementarse como tipo SQL de muchas formas distintas, no siendo estas compatibles entre sí. Es ahí donde hacen su aparición los estándares, los cuales veremos con más detalle en el capítulo 24.7.

De especial importancia en este sentido es la norma denominada *Simple Features for SQL*, que especifica cómo han de implementarse los tipos SQL correspondientes a los objetos espaciales, con objeto de estandarizar esta implementación. Así, una base de datos incorporará dichos tipos en sus modelos y estos podrán ser posteriormente utilizados para la realización de consultas SQL. En el ya citado capítulo 24.7 explicaremos en detalle este y otros estándares.

El aspecto más importante para el contenido de este apartado no es el modelo de datos u otras características de la base de datos en sí (vimos esto ya con más detalle en el capítulo 8), sino la forma en que esto afecta a la realización de consultas. Por ello, la parte de mayor interés son los métodos que esos tipos implementan, y que pueden emplearse para dar forma a consultas espaciales como las que veíamos en el apartado previo.

Estos métodos vienen también especificados por la norma Simple Features, y como ya veremos podemos dividirlos en tres grupos: funciones básicas, operaciones topológicas y de conjunto, y operaciones de análisis espacial. Los resultados que arrojan estos métodos se pueden emplear para dar forma a consultas que realizan operaciones como la selección o la unión. Ya vimos cómo llevar estas a cabo mediante consultas SQL, y los ejemplos con contenido espacial del apartado son también operaciones de este tipo, bien sean consultas o uniones. Veamos, pues, cómo podrían realizarse mediante consultas SQL empleando los métodos que han de presentar los tipos que cumplen la especificación Simple Features.

Por ejemplo, para ver qué países atraviesa el río Nilo, podemos emplear una consulta como la siguiente:

```
SELECT Paises.Nombre,
FROM Rios, Paises
WHERE Cross(Rios.shape, Paises.Shape)
      AND Rios.Nombre = 'Nilo'
```

La expresión `Cross(Rios.Shape, Paises.Shape)` hace uso del método `Cross`, que devuelve 1 en caso de que las dos geometrías pasadas como parámetros se intersequen, y 0 en caso contrario. Este se utiliza para realizar la selección solo sobre aquellas que cumplan la condición de ser cortadas por una geometría dada, en este caso la del río Nilo.

La relación espacial entre el río y los distintos países no puede evaluarse haciendo uso de SQL sin extensiones espaciales, puesto que sin ellas no es posible interpretar correctamente las geometrías almacenadas en la base de datos y proceder con ellas a calcular las correspondientes intersecciones.

Algunos métodos como el método `Cross` anterior expresan condiciones, y al evaluarse devuelven valores 1 o 0 (verdadero/falso). Otros dan como resultados tipos distintos de valores, entre los que pueden estar nuevas geometrías.

Un ejemplo de esto podemos verlo al combinar el método `Length`, que devuelve un valor numérico correspondiente a la longitud de una línea, con el método `Intersection`, que devuelve una nueva geometría a partir de la intersección de otras dos. Mediante ambos, podemos resolver la consulta anterior pero añadiendo en la tabla resultado no solo el nombre de los países que son

atravesados por el Nilo, sino también la distancia que este río recorre a través de cada uno de ellos.

La consulta tendría en este caso una forma como la siguiente:

```
SELECT Paises.Nombre,
      Length(Intersection(
             Rios.Shape, Paises.Shape))
       AS 'Longitud'
FROM Rios, Paises
WHERE Cross(Rios.Shape, Paises.Shape)
      AND Rios.Nombre = 'Nilo'
```

Al igual que en el apartado anterior, el objetivo de este apartado no es mostrar con detalle la sintaxis del lenguaje SQL cuando este se emplea para la realización de consultas espaciales. Los anteriores son únicamente algunos ejemplos para poner de manifiesto la potencia de este planteamiento y mostrar cómo los elementos espaciales se integran en el lenguaje SQL. Puede encontrarse más información en [106].

Además de esta extensión a SQL, existen otras propuestas alternativas tales como Geo–SQL, SSQL (Spatial SQL) o SQL-MM.

12.4.2. Índices espaciales

Si realizamos una consulta a una base de datos, el resultado es un subconjunto de esta con los elementos que cumplen el criterio expresado en la consulta. Si se implementa de forma *directa* dicha consulta, esta operación implica comprobar todos los elementos de la base de datos y ver cuáles son los que cumplen con el citado criterio. Teniendo en cuenta que una base de datos puede tener un gran tamaño, esta forma de proceder no es la óptima.

Veamos un ejemplo para poder entender mejor esto. Supongamos que tenemos una guía telefónica, que no es sino una base de datos en la que cada registro contiene dos campos: nombre y apellidos, y teléfono. ¿Cómo buscaríamos en esa guía telefónica el numero de una persona llamada Juan Pérez? Sin duda, leyendo uno por uno todos los nombres acabaríamos encontrando el que buscamos y su número correspondiente, pero antes tendríamos que leer una gran cantidad de nombres y apellidos (más aún en este caso, considerando que la letra P se encuentra en la mitad final del alfabeto), con lo que no resulta una opción muy lógica. En tal caso, una guía telefónica sería una herramienta inútil.

Sin embargo, habitualmente consultamos guías telefónicas sin problemas y encontramos rápidamente el teléfono de una persona sin necesidad de leer más que unos pocos nombres. Esto es así porque sabemos cómo están dispuestos los datos y buscando en el índice sabemos incluso en qué página comienzan los apellidos con una letra dada (en este caso la letra P). El uso de este índice nos permite optimizar el proceso de búsqueda de una forma realmente radical.

Al utilizar una base de datos, si no disponemos de un índice deberemos recorrer toda ella para dar respuesta a nuestras consultas. No sabemos *dónde* buscar las respuestas a nuestras consultas, del mismo modo que si en una guía telefónica no supiéramos que carece de sentido buscar en la letra F el número telefónico del señor Pérez.

Los índices nos permiten *alcanzar* los elementos que constituyen la respuesta a nuestra consulta, haciéndolo de la forma más rápida y llegando hasta ellos sin tener que pasar por todos los restantes.

Describir los índices empleados en bases de datos no espaciales requiere describir asimismo estructuras de datos complejas que escapan del alcance de este texto (los denominados *árboles B+* son las estructuras utilizadas con más frecuencia para esta tarea). Por esta razón, no se detallarán en este capítulo más allá de la anterior descripción básica, pudiéndose encontrar más informa-

ción en las referencias proporcionadas a lo largo del capítulo.

Más interesantes que estos índices nos resultan aquellos que se utilizan en las bases de datos espaciales, que denominamos *índices espaciales*. El concepto es similar al de índices de bases de datos no espaciales: elementos que permiten optimizar las consultas mediante una correcta estructuración de los datos, en particular en este caso de su componente espacial.

Los índices espaciales no deben resultarnos desconocidos, ya que los vimos en el capítulo 5, estudiando en su momento los tipos existentes y su proceso de creación. Ahora en este capítulo veremos el verdadero uso práctico de estos, y así podremos comprender mejor la necesidad de su existencia.

Puede entenderse igualmente la idea de un índice espacial mediante un sencillo ejemplo de cómo empleamos ideas parecidas a los índices espaciales de forma natural cuando tratamos de resolver una consulta espacial sin la ayuda de un SIG. Supongamos que tenemos nuestro mapa de países del mundo y queremos averiguar qué países tienen su frontera a menos de 2000 kilómetros de la frontera de España. ¿Cómo operaríamos de manera natural para dar respuesta a esta consulta?

La solución más inmediata es medir la distancia entre España y todos los países restantes, y después tomar aquellos que hayan arrojado un resultado de distancia menor a 2000. La operación daría el resultado esperado, pero implicaría un gran número de mediciones, y no sería una forma óptima de operar. De hecho, es probable que a nadie se le ocurriese operar de esta forma en ningún caso. Por ejemplo, lo más probable es que no efectuemos mediciones con los países de América, pues un conocimiento básico de geografía basta para saber que todos ellos se encuentran a más de 2000 kilómetros. No sabemos exactamente a qué distancia se encuentran, pero sabemos que de ningún modo van a cumplir el criterio establecido en la consulta. De modo similar podemos eliminar Australia y gran parte de Asia, porque se encuentran en una situación similar.

Ese conocimiento básico de geografía que tenemos es en realidad una especie de índice espacial. No sirve para saber las distancias exactas ni resolver la consulta por completo, pero sirve para dar una aproximación y facilitar el trabajo. Descartamos un buen numero de países de forma casi inmediata, y luego solo realizamos las operaciones costosas (la medición) con un subconjunto del total. En nuestra mente, tenemos el conocimiento estructurado a distintos niveles. Incluso si memorizamos todas esa distancias, existe otro nivel más general de conocimiento, a otra escala, siendo este el que nos indica de forma rápida que toda América está fuera de la distancia establecida en la consulta y no merece la pena efectuar mediciones referidas a países de ese continente.

Con la utilización un índice espacial, el proceso de consulta espacial se compone de dos subprocesos: *filtrado* y *refinamiento* [107]. En el proceso de filtrado se hace una primera selección aproximada de entidades, las cuales son candidatas a cumplir los criterios de la consulta. Se reduce de este modo el número de elementos sobre los que se ha de trabajar, y esta reducción, apoyada en los índices espaciales, tiene un coste operacional menor que aplicar la consulta en sí a todos los elementos.

En el refinamiento, se toman los elementos que han superado la fase de filtrado, y sobre ellos se aplica la consulta como tal. Esto tendrá como consecuencia que algunos de estos elementos, pese a haber pasado la primera fase de filtrado, no cumplan el criterio de la consulta, ya que este filtrado era una aproximación al resultado final. De esta forma refinamos este resultado previo y

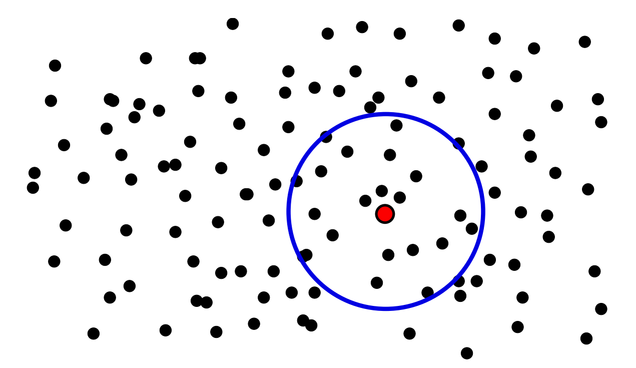


Figura 12.4: Para calcular los diez puntos más cercanos a una coordenada dada (en rojo), nuestra intuición espacial nos ayuda a decidir que estos se han de buscar en un subconjunto reducido tal como el de los situados dentro del circulo azul de la figura. Este proceso de *filtrado* y *refinamiento* ahorra operaciones, y es el fundamento conceptual de los índices espaciales.

obtenemos ya la respuesta exacta a la consulta formulada.

En resumen, se puede decir que los índices espaciales nos permiten obtener resultados en un área concreta sin necesidad de analizar todo el espacio ocupado por el total de los datos. Estos índices espaciales no son exclusivos del trabajo con bases de datos y la realización de consultas espaciales, sino que se encuentran implícitos en muchas operaciones que vamos a ver en los próximos capítulos dentro de esta parte del libro. Estas operaciones en realidad necesitan para su desarrollo efectuar algún tipo de consulta, y dicha consulta depende de los índices espaciales para ejecutarse con un buen rendimiento.

Por ejemplo, las funciones de interpolación, que veremos en el capítulo 14, para calcular el valor en una coordenada concreta, y a partir de los valores de una capa de puntos, habitualmente utilizan los n puntos más cercanos a dicha coordenada. Para saber cuáles son estos n puntos, podrían calcularse las distancias desde todos los puntos de la capa hasta la coordenada en cuestión, y después tomar los n para los cuales esa distancia es menor. Esta forma de proceder, sin embargo, requiere un número de cálculos demasiado elevado, que haría imposible ejecutar en un tiempo lógico dichos algoritmos de interpolación cuando los puntos de la capa sean numerosos (lo cual es muy frecuente).

Si vemos la figura 12.4, para calcular cuales son los diez puntos (en negro) más cercanos a una coordenada dada (en rojo), no mediríamos las distancias de todos ellos. Mirando a simple vista podemos estimar que esos puntos van a estar dentro de un círculo aproximadamente como el representado en la figura, y podemos prescindir de los restantes a la hora de calcular las distancias exactas. Dentro de ese circulo hay más de diez puntos, con lo cual debe *refinarse* ese resultado antes de poder ofrecer una respuesta exacta a la consulta.

Otros procesos en los que son vitales los índices espaciales son las operaciones de solape entre capas de polígonos, que veremos en el capítulo 19. Sin ellos, el rendimiento de estas operaciones espaciales sería mucho menor o incluso, como en el caso de la interpolación, totalmente insuficiente para que tales operaciones se puedan aplicar en la mayoría de los casos.

12.5. Resumen

Las consultas son uno de los análisis fundamentales dentro de un SIG. Básicamente, una consulta efectúa una pregunta acerca de la información contenida en una capa, y obtiene como resultado los elementos de la capa que dan respuesta a dicha pregunta. Las consultas son en general un elemento aplicado sobre capas vectoriales, y el resultado de la consulta se expresa mediante una selección de entidades dentro de aquellas que componen dicha capa.

Las consultas pueden efectuarse sobre la componente temática del dato geográfico, en cuyo caso emplean los mismos mecanismos que las bases de datos fuera de un SIG. Esto incluye el empleo de lenguajes de consulta, específicamente desarrollados para esta tarea. El lenguaje SQL (Structured Query Language) es el más habitual de estos.

Al incorporar la componente espacial, se añaden nuevos elementos para realizar consultas. Los criterios de consulta añaden predicados espaciales basados en las relaciones entre las distintas entidades, y estos a su vez pueden combinarse con los predicados no espaciales para la formulación de consultas complejas. Los lenguajes de consulta se extienden para dar cabida a estos nuevos predicados, así como a funciones espaciales basadas en las propias entidades de las capas.

La aplicación de criterios espaciales hace necesaria la utilización de índices espaciales para optimizar el trabajo con grandes volúmenes de datos. Estos índices estructuran los datos de tal modo que en la realización de consultas espaciales no es necesario efectuar dicha consulta sobre la totalidad de los datos, sino únicamente sobre una fracción de ellos.

Capítulo 13

Estadísticas espaciales

13.1. Introducción

La información espacial es susceptible de ser analizada estadísticamente como cualquier otro tipo de información. Una serie de n datos recogidos en otros tantos puntos no deja de ser una serie de datos sobre la que pueden aplicarse las técnicas estadísticas habituales. Junto a ello, cada uno de estos datos tiene asociada una coordenada, y esta aporta una información adicional que puede emplearse igualmente para obtener resultados estadísticos de diversa índole.

Si trabajamos en el plano cartesiano, en lugar de una serie de valores de una variable a disponemos de una serie de ternas (x, y, a). Extendiendo la posibilidad de analizar estadísticamente los valores a recogidos en esa serie de localizaciones, encontramos otras dos formas de analizar este conjunto.

- Analizar la disposición espacial, con independencia de los valores. Es decir, estudiar el conjunto de pares de valores (x, y)
- Analizar la disposición espacial y los valores recogidos. Es decir, estudiar el conjunto de ternas (x, y, a)

Este tipo de análisis se lleva a cabo preferentemente sobre capas de tipo punto, aunque algunas de estas formulaciones pueden igualmente aplicarse a capas ráster, considerando que cada celda conforma de igual modo una terna de valores.

Dentro de este capítulo analizaremos algunos de los analisis estadísticos espaciales más básicos, como son los siguientes:

- **Medidas centrográficas**. El equivalente espacial de las medidas de tendencia central como el momento de primer orden (media) o la mediana, así como de las de dispersión tales como el momento de segundo orden (desviación típica).
- **Análisis estadístico de líneas**. Descriptores estadísticos para líneas y ángulos.
- **Análisis de patrones de puntos**. Este tipo de análisis permite caracterizar la estructura espacial de un conjunto de puntos en función de parámetros como la densidad o las distancias entre puntos y su configuración en el espacio.
- **Autocorrelación espacial**. Los puntos cercanos tienden a tener valores más similares entre sí que los puntos alejados. Este fenómeno puede cuantificarse y estudiarse con una serie de índices, así como mediante elementos tales como variogramas o correlogramas.

Otro grupo de procedimientos con componente estadística, los relativos a las técnicas de interpolación, se verán en un capítulo independiente (Capítulo 14). De igual modo, aquellos que permiten el calculo de densidades se recogen también en dicho capítulo.

Un capítulo dedicado a la aplicación de otras técnicas estadísticas más complejas tales como técnicas de agrupación o regresiones espaciales

(Capitulo 21) completa este grupo de secciones dedicadas a los elementos estadísticos.

13.2. Medidas centrográficas

Las medidas centrográficas representan descriptores básicos de los datos espaciales, extendiendo las medidas de tendencia central y dispersión de la estadística clásica al ámbito espacial.

La principal medida de tendencia central espacial es el *centro medio*. El centro medio es un punto cuyas coordenadas son la media en cada eje de las coordenadas de los puntos analizados (Figura 13.2). Es decir, el punto $(\overline{x}, \overline{y})$ tal que

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{N} \tag{13.2.1}$$
$$\overline{y} = \frac{\sum_{i=1}^{n} y_i}{N}$$

Cada uno de los puntos puede ponderarse según el valor recogido en él, de forma que lo anterior quedaría como

$$\overline{x} = \frac{\sum_{i=1}^{N} a_i x_i}{\sum_{i=1}^{N} a_i} \tag{13.2.2}$$
$$\overline{y} = \frac{\sum_{i=1}^{N} a_i y_i}{\sum_{i=1}^{N} a_i}$$

El centro medio es el centro de gravedad del conjunto de puntos, tomando como masa de cada uno el valor asociado a este. Asimismo, es el punto que minimiza la suma de distancias al cuadrado, esto es, la expresión

$$\sum_{i=1}^{N} d_{ic} = \sum_{i=1}^{N} (\overline{x} - x_i)^2 + (\overline{y} - y_i)^2 \tag{13.2.3}$$

Un uso habitual del centro medio lo encontramos en los estudios demográficos, que pueden analizar la evolución de las poblaciones sobre el territorio estudiando cómo se ha desplazado su centro medio a través del tiempo.

El equivalente espacial de la mediana es el *centro mediano*. Al igual que el centro medio, el centro mediano es también un punto. En este caso sus coordenadas son las medianas de las de los puntos analizados en cada eje [108].

Puede ser interesante también analizar el centro mediano como una línea en lugar de un punto. Por ejemplo, una línea vertical que pasa por la componente en x del centro mediano. Si trabajamos con una serie de puntos que representan poblaciones y estos se ponderan según su número de habitantes, esta línea divide el territorio en dos zonas igualmente pobladas. La mitad de los habitantes viven a un lado de ella, y la otra mitad al otro lado.

En la figura 13.1 puede verse una representación de lo anterior.

El inconveniente del centro mediano es que depende de los ejes escogidos y no es por tanto invariante ante rotaciones. Para solucionar esto suele emplearse como definición alternativa la de aquel punto del espacio que hace mínima la suma de distancias a todos los puntos de datos[109]. El cálculo de este punto requiere de un proceso iterativo [110] en el cual se tiene que

$$x = \frac{\sum_{i=1}^{N} \frac{d_i}{a_i} x_i}{\sum_{i=1}^{N} \frac{d_i}{a_i}} \tag{13.2.4}$$
$$y = \frac{\sum_{i=1}^{N} \frac{d_i}{a_i} y_i}{\sum_{i=1}^{N} \frac{d_i}{a_i}} \tag{13.2.5}$$

donde d_i es la distancia del punto i-ésimo a la localización del centro mediano en la iteración actual. Como primera coordenada para iterar, una buena elección es el propio centro medio. El proceso se detiene cuando la distancia entre el nuevo

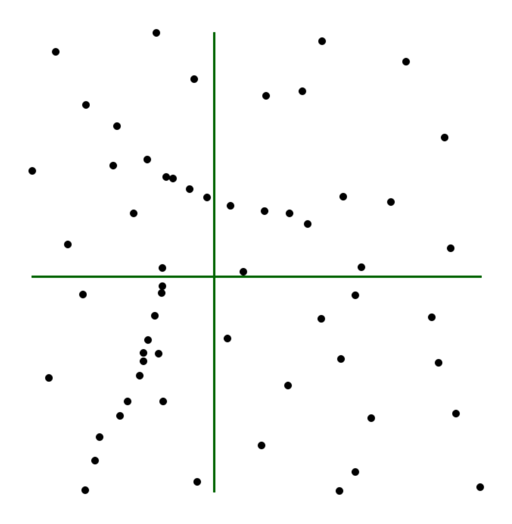


Figura 13.1: Centro mediano y líneas de división pasando por este, las cuales dividen el conjunto de puntos en dos partes iguales a Este y Oeste, y Norte y Sur, respectivamente.

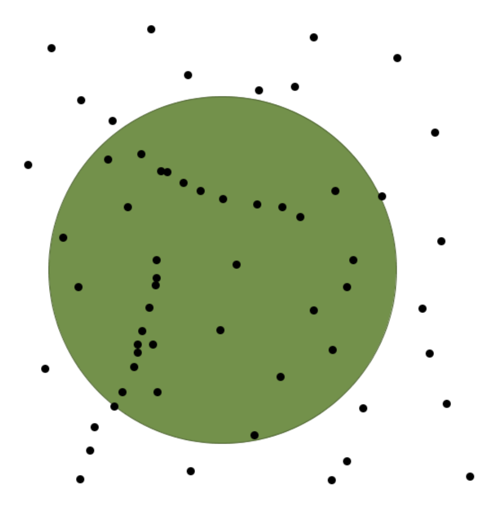


Figura 13.2: Circulo de radio igual a la distancia típica centrado en el centro medio del conjunto.

centro mediano y el de la iteración anterior es menor que un determinado umbral establecido de antemano.

Respecto a las medidas de dispersión, el equivalente a la desviación típica es la denominada *distancia típica*, cuya expresión es la siguiente [111]

$$s_d = \sqrt{\frac{\sum_{i=1}^{n} d_i^2}{n}} \tag{13.2.6}$$

siendo d_i la distancia entre el punto i-ésimo y el centro medio.

También puede escribirse lo anterior como

$$s_d = \sqrt{\left(\frac{\sum_{i=1}^{N} x_i^2}{N} - \overline{x}^2\right) + \left(\frac{\sum_{i=1}^{N} y_i^2}{N} - \overline{y}^2\right)} \tag{13.2.7}$$

Es interesante comentar que la distancia a la media en el concepto habitual de desviación típica puede ser positiva o negativa (de ahí que se eleve al cuadrado y después se aplique la raíz), mientras que en el caso espacial es siempre positiva.

Una forma de representar esta distancia típica es mediante un circulo de radio dicha distancia centrado en el centro medio (Figura 13.2)

La distancia típica puede, igualmente, calcularse ponderando los distintos puntos, quedando su expresión como

$$s_d = \sqrt{\frac{\sum_{i=1}^{n} a_i d_i^2}{\sum_{i=1}^{N} a_i}} \tag{13.2.8}$$

Mediante esta representación se asume, no obstante, que la dispersión es la misma en todas direcciones. Esta simplificación raramente es cierta, y es más correcto definir en lugar de un círculo una *elipse de desviación*. Esta elipse de desviación viene definida por sus semiejes mayor y menor, en los cuales se dan, respectivamente, la mayor y menor dispersión.

El ángulo α que define al semieje mayor x' viene expresado según

$$\tan\alpha = \frac{\sum_{i=1}^{N} dx_i - \sum_{i=1}^{N} dy_i}{2\sum_{i=1}^{N} dx_i dy_i} + \frac{\sqrt{\left(\sum_{i=1}^{N} dx_i - \sum_{i=1}^{N} dy_i\right)^2 + 4\sum_{i=1}^{N} dx_i dy_i}}{2\sum_{i=1}^{N} dx_i dy_i}$$

siendo dx_i y dy_i las distancias en los ejes x e y respectivamente entre el punto i–ésimo y el centro medio.

El semieje menor es perpendicular al anterior.

Las distancias típicas en cada uno de estos dos semiejes vienen expresadas por

$$\delta_{x'} = \left(\sum_{i=1}^{N} dx_i^2 \cos^2\alpha + 2\left(\sum_{i=1}^{N} dx_i dy_i\right)\sin\alpha\cos\alpha + \sum_{i=1}^{N} dy_i^2 \sin^2\alpha\right)^{\frac{1}{2}} \qquad (13.2.9)$$

$$\delta_{y'} = \left(\sum_{i=1}^{N} dx_i^2 \sin^2\alpha + 2\left(\sum_{i=1}^{N} dx_i dy_i\right)\sin\alpha\cos\alpha + \sum_{i=1}^{N} dy_i^2 \cos^2\alpha\right)^{\frac{1}{2}} \qquad (13.2.10)$$

Por último, la medida de desviación relativa que equivale en la estadística espacial al coeficiente de variación es la *distancia relativa* [112], que se calcula dividiendo la distancia típica por el radio de un círculo con el mismo área que la zona de estudio. Si esta zona es circular, se tiene por tanto

$$s_{d,rel} = \frac{s_d}{R} \qquad (13.2.11)$$

siendo R el radio de la zona de estudio.

En caso de que esta zona sea cuadrada y de área A , se tiene que

$$s_{d,rel} = \frac{s_d\sqrt{\pi}}{\sqrt{A}} \qquad (13.2.12)$$

En la figura 13.3 puede verse cómo distribuciones espaciales iguales (con la misma distancia típica) representan dispersiones relativas distintas en relación a la zona de estudio.

13.3. Estadísticas sobre líneas. Variables circulares

Dentro de los objetos geográficos, las líneas merecen algunos comentarios aparte en lo que a su análisis respecta. Tanto las líneas como los polígonos pueden ser reducidos en ultima instancia a puntos (los polígonos bien por sus puntos constituyentes o bien por el centroide, el cual coincide con el centro medio), y analizados estos con algunas de las fórmulas antes vistas o las que se verán más adelante. La particularidad de las líneas estriba en que, además de valores puntuales o de área (como los de los polígonos), definen igualmente direcciones y ángulos de giro entre sus segmentos. El análisis estadístico de variables circulares como estas presenta sus propias particularidades, que deben conocerse para poder extraer resultados correctos a partir de datos de esta índole.

Un ejemplo del uso de variables direccionales lo encontramos, por ejemplo, en el estudio de desplazamientos de animales cuyas rutas hayan sido monitorizadas y se encuentren dentro de un SIG como capas de líneas. Un situación similar se

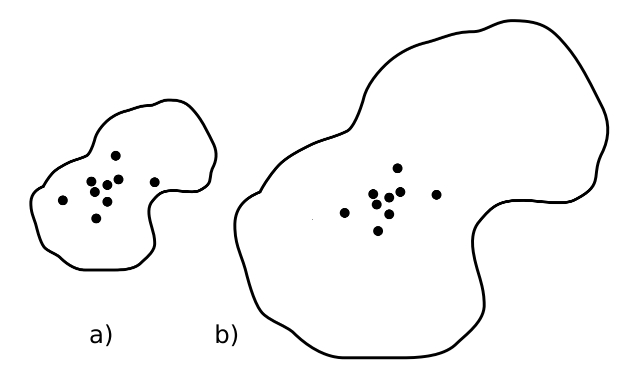


Figura 13.3: Aunque ambas distribuciones tienen la misma distancia típica, la dispersión en relación al área analizada es distinta, con lo que los valores de distancia relativa serán distintos.

da en el caso de elementos que no representen un movimiento pero tengan dirección, tales como fallas u otros elementos geológicos. Más alla de casos como estos, los conceptos relativos a este tipo de variables también tienen aplicación para cualquier información similar, con independencia de su formato de almacenamiento. Así, son de aplicación, entre otros, para el estudio de orientaciones dentro del análisis geomorfométrico (Capítulo 16), el cual se lleva a cabo fundamentalmente sobre capas ráster.

En el caso que nos ocupa del estudio de líneas, pueden considerarse todos y cada uno de los segmentos de estas como líneas en sí, o bien la línea ficticia que une el inicio del primer segmento con el final del último.

A continuación se mostrarán brevemente los estadísticos más frecuentes para datos circulares, con especial énfasis en su aplicación al análisis de líneas dentro de un SIG. Descripciones más detalladas de estos y otros elementos de estadística circular, junto a sus aplicaciones en áreas donde el empleo de SIG es habitual, pueden consultarse en [113] o [114].

Para comenzar, el cálculo de la media de dos ángulos ejemplifica bien las particularidades de los datos circulares. Sean tres ángulos de 5°, 10°y 15°respectivamente. El concepto habitual de media aplicado a estos valores resultaría en un ángulo medio de 10°, correcto en este caso. Si giramos ese conjunto de ángulos 10 grados en sentido antihorario, dejándolos como 355°, 0°, 5°, la media debería ser 0°, pero en su lugar se tiene un valor medio de 120°.

Una forma correcta de operar con ángulos $\alpha_1, \ldots, \alpha_n$ consiste en hacerlo con las proyecciones del vector unitario según dichos ángulos, es decir $\sin\alpha_1, \ldots. \sin\alpha_n$ y $\cos\alpha_1, \ldots. \cos\alpha_n$. Aplicando luego los estadísticos habituales sobre estos valores se obtienen unos nuevos valores de senos y cosenos que permiten obtener el ángulo resultante aplicando sobre ellos la función arcotangente.

En el caso de segmentos orientados tales como los que constituyen las líneas dentro de una capa de un SIG, resulta conveniente tratar cada segmento como un vector. La resultante de su suma vectorial será otro vector con la dirección media de todos los segmentos, y cuyo módulo (longitud) aporta información acerca de la tendencia y variación de las direcciones a lo largo de la linea. Si la dirección es uniforme, el módulo será mayor, siendo menor si no lo es (Figura 13.4). El vector resultante puede dividirse por el número total de segmentos iniciales para obtener una media vectorial.

Es decir, se tiene un vector cuya orientación viene definida por un ángulo $\overline{\alpha}$ tal que

$$\overline{\alpha} = \arctan \frac{S}{C} \tag{13.3.1}$$

y con un módulo $\overline{R}$ según

$$\overline{R} = \frac{\sqrt{S^2 + C^2}}{N} \tag{13.3.2}$$

siendo S y C las sumas de senos y cosenos, respectivamente.

$$S = \sum_{i=1}^{N} \sin \alpha_i \quad ; \quad S = \sum_{i=1}^{N} \cos \alpha_i \tag{13.3.3}$$

El módulo $\overline{R}$ se conoce también como *concentración angular* y es una medida inversa de la dispersión angular. No obstante, hay que tener en cuenta que valores próximos a cero, los cuales indicarían gran dispersión, puede proceder de dos agrupaciones de ángulos similares (es decir, con poca dispersión) si estas agrupaciones se diferencian entre sí 180°.

Cuando se trabaja con direcciones en lugar de orientaciones, es frecuente multiplicar por dos los valores angulares y posteriormente simplificar el ángulo aplicando módulo 360°. Es decir, aplicar la transformación $\alpha' = 2\alpha \mod 360°$.

La forma en que las distintas orientaciones se congregan en torno a la media, relacionada directamente con la dispersión, puede servir para inferir la existencia de una dirección predominante o bien que los valores angulares se hallan uniformemente distribuidos. La comprobación de que existe una tendencia direccional es de interés para el estudio de muchos procesos tales como el estudio de movimiento de individuos de una especie, que puede denotar la existencia de una línea migratoria preferida o revelar la presencia de algún factor que causa dicha predominancia en las direcciones.

Existen diversos test que permiten aceptar o rechazar la hipótesis de existencia de uniformidad, entre los cuales destacan el test de Rayleigh, el test V de Kuiper [115] o el test de espaciamiento de Rao [116]

Para este último, se tiene un estadístico U según

$$U = \frac{1}{2} \sum_{i=1}^{N} \|T_i - \lambda\| \tag{13.3.4}$$

siendo

$$\lambda = \frac{360}{N} \tag{13.3.5}$$

$$T_i = \begin{cases} \alpha_{i+1} - \alpha_i & \text{si } 1 \leq i < N-1 \\ 360 - \alpha_n + \alpha_1 & \text{si } i = N \end{cases} \tag{13.3.6}$$

Puesto que las desviaciones positivas deben ser iguales a las negativas, lo anterior puede simplificarse como

$$U = \sum_{i=1}^{N} (T_i - \lambda) \tag{13.3.7}$$

Para un numero de puntos dado y un intervalo de confianza establecido, los valores de U están tabulados, y pueden así rechazarse o aceptarse la hipótesis nula de uniformidad. Dichas tablas pueden encontrarse, por ejemplo, en [117].

13.4. Análisis de patrones de puntos

Las coordenadas de un conjunto de puntos no solo representan una información individual de cada uno de ellos, sino de igual modo para todo el conjunto a través de las relaciones entre ellas. La disposición de una serie de puntos en el

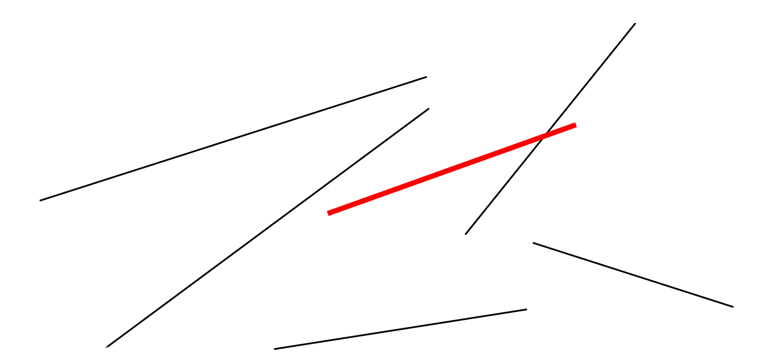


Figura 13.4: Media vectorial (en rojo) de una serie de segmentos.

espacio conforma lo que se conoce como un *patrón de puntos*, el cual puede aportar información muy valiosa acerca de las variables y procesos recogidos en dichos puntos. Por ejemplo, si estos representan lugares donde se han observado individuos de una especie, su distribución espacial puede servir como indicador de la interacción entre dichos individuos o con el medio.

La caracterización de un patrón de puntos es, por tanto, de interés para la descripción de estos, y se realiza a través de análisis estadísticos y descriptores que definen la estructura del mismo.

Para llevar a cabo este análisis se asume que la estructura espacial de un patrón dado es el resultado de un *proceso puntual*. Se entiende por proceso puntual un proceso estocástico que genera tales patrones, compartiendo todos ellos una similar estructura (la ley de dicho proceso). Los puntos son eventos de dicho proceso. Describiendo el tipo de patrón se obtiene información sobre el proceso puntual que lo ha originado.

Podemos encontrar múltiples ejemplos de procesos puntuales, tales como la disposición de individuos de una especie, la disposición de los árboles en un bosque o la aparición de casos de una enfermedad. Cada uno de ellos tiene sus propias características.

Como se puede observar en la figura 13.5, existen tres tipos de patrones que un proceso de puntos puede generar:

- **Agregado**. La densidad de los puntos es muy elevada en ciertas zonas.
- **Aleatorio**. Sin ninguna estructura, las posiciones de los puntos son independientes entre sí.
- **Regular**. La densidad es constante y los puntos se disponen alejados entre sí.

El análisis de patrones de puntos se fundamenta básicamente en la comparación entre las propiedades de una distribución teórica aleatoria (distribución de Poisson) y las de la distribución observada. La distribución teórica aleatoria cumple que se da *aleatoriedad espacial completa* (CSR, *Complete Spatial Randomness*, en inglés). De este modo, se puede decidir si la distribución observada es también aleatoria en caso de existir similitud, o bien es de alguno de los dos tipos restantes, según sea la discrepancia existente.

Las propiedades a comparar pueden ser:

- **Propiedades de primer orden**. La intensidad del proceso $\lambda(h)$, definida como la densidad (número de puntos por unidad de área). En general, se asume que es una propiedad estacionaria, esto es, constante a lo largo de la zona de estudio. Existen distribuciones como la *distribución no*

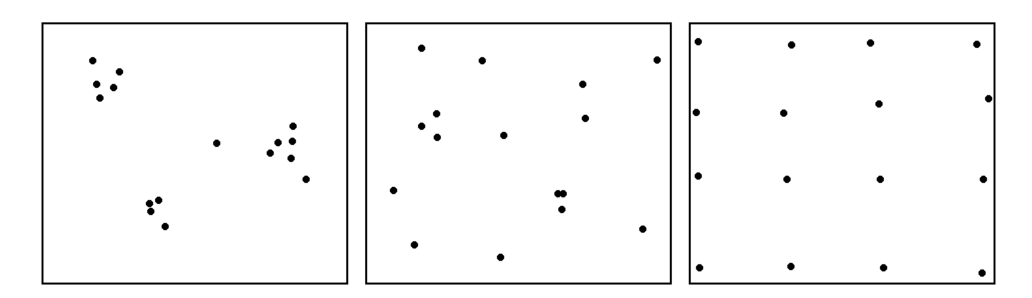


Figura 13.5: De izquierda a derecha, patrones de puntos agregado, aleatorio y regular.

homogénea de Poisson que asumen una variabilidad de la intensidad a lo largo de la zona de estudio. En el apartado 14.3 veremos cómo crear capas continuas de esta intensidad $\lambda(h)$.

- **Distancia entre puntos**. Relaciones entre cada punto con los de su entorno. Basado en las denominadas *propiedades de segundo orden.*

13.4.1. Análisis de cuadrantes

En el primero de los casos, la metodología de *análisis de cuadrantes* divide la zona de estudio en unidades regulares, *cuadrantes*, y estudia el número de puntos que aparecen dentro de cada una.

La forma de estas unidades puede ser cualquiera, aunque lo habitual es emplear unidades cuadradas, de ahí la denominación. Debido a los efectos de escala, el tamaño de estas unidades tiene una gran influencia en los resultados obtenidos. Un tamaño habitual es el doble del área media disponible para cada punto, es decir, cuadrados cuyo lado tendrá una longitud

$$l = \sqrt{\frac{2A}{N}} \quad (13.4.1)$$

siendo N el número de puntos y A el área de la zona de estudio.

Con la serie de datos que indica el conteo de puntos en cada cuadrante, se procede al análisis estadístico. Este puede hacerse comparando los conteos en los cuadrantes o según la relación entre la media y la varianza de la serie. En este segundo caso, partimos de que en una distribución aleatoria es de esperar una varianza igual a la media [118]. Por tanto, el cociente entre la varianza y la media debe ser cercano a 1. Si en la distribución analizada este cociente está próximo a ese valor, se tratará de una distribución aleatoria. En una distribución uniforme, la varianza (y por tanto el cociente con la media) será cercana a 0. En las distribución agrupadas, la varianza sera mayor, y el cociente por tanto superior a 1.

El análisis de cuadrantes no es en realidad una medida del patrón, sino de la dispersión. Además, debido al uso de una unidad de análisis (el cuadrante) fija, puede no ser capaz de localizar agrupamientos locales en esta.

Otra debilidad de este método es que no es capaz de diferenciar entre distribuciones tales como las de la figura 13.6, claramente distintas pero que arrojan un resultado idéntico al aplicar esta metodología con los cuadrantes mostrados.

No obstante, la aplicación de este método en campos como la biología es muy habitual, y se han desarrollado numerosas extensiones del mismo tales como el *índice de David–Moore* [119], el *índice de frecuencia de agregados* [120], o el índice I_δ de [121], entre otros muchos.

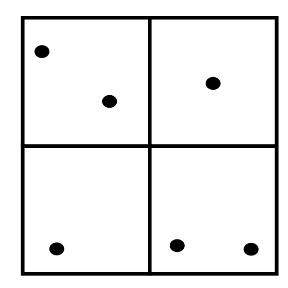
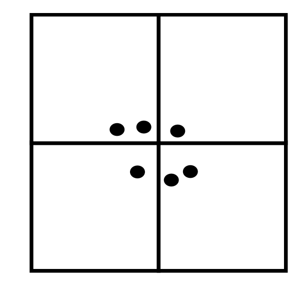



Figura 13.6: Dos disposiciones de puntos distintas que darían un mismo resultado al analizarse por el método de cuadrantes.

13.4.2. Análisis de vecino más cercano

El *método de vecino más cercano* [122] permite solventar algunos de los problemas asociados al análisis de cuadrantes. Para ello, se basa en las distancias de cada punto a su vecino más cercano. Comparando estas distancias con el valor que cabe esperar en una distribución aleatoria, puede deducirse el tipo de estructura en la distribución observada.

El valor que define el patrón de puntos a estudiar es el *índice de vecino más cercano*, que se calcula como

$$I_{mc} = \frac{\overline{d}_{mc}}{E(\overline{d}_{mc})} \qquad (13.4.2)$$

siendo $\overline{d}_{mc}$ la media de las distancias al punto más cercano, según

$$\overline{d}_{mc} = \frac{\sum_{i=1}^{N} d_{mc}}{N} \qquad (13.4.3)$$

$E(\overline{d}_{mc})$ es la media esperada en una distribución de Poisson, y se calcula según la expresión

$$\hat{\mu} = \frac{1}{2\sqrt{\lambda}} \qquad (13.4.4)$$

siendo λ la densidad de puntos por unidad de área, es decir

$$\lambda = \frac{N}{A} \qquad (13.4.5)$$

[123] propone corregir lo anterior para tener en cuenta los efectos de borde, utilizando la siguiente expresión:

$$\hat{\mu} = \frac{1}{2\sqrt{\lambda}} + 0{,}0514 + \frac{0{,}041}{\sqrt{N}}\frac{B}{N} \qquad (13.4.6)$$

donde B es la longitud del perímetro del área estudiada.

El índice de vecino más cercano tiene un valor de 1 en una distribución aleatoria, menor de 1 en una distribución agregada, y mayor en una regular.

La desviación típica de las distancias se estima según

$$\hat{\sigma}_d = \sqrt{\frac{4-\pi}{4\pi\frac{N^2}{A}}} \qquad (13.4.7)$$

Aplicando como en el caso de la media una corrección de los efectos de borde, se tiene

$$\hat{\sigma}_d = \sqrt{0{,}070\frac{A}{N^2} + 0{,}037B\sqrt{\frac{A}{N^5}}} \qquad (13.4.8)$$

Conociendo este resultado y que bajo la hipótesis de aleatoriedad espacial completa puede asumirse una distribución normal de los valores de distancia con la media y la desviación típica anteriores, pueden hacerse test de significación

para conocer con qué grado de confianza es posible afirmar que la distribución analizada es o no aleatoria.

El análisis de vecino más cercano puede ampliarse al de los n vecinos más cercanos. No obstante, este tipo de formulaciones se implementan con mucha menor frecuencia y son significativamente más complejas que las basadas en un único punto vecino.

13.4.3. Función K de Ripley

El problema de escala vimos que era patente en el método del análisis de cuadrantes, puesto que existía una fuerte dependencia del tamaño del cuadrante. La función K de Ripley trata de incorporar la escala como una variable más del análisis, convirtiendo dicha dependencia en un hecho favorable en lugar de una desventaja.

Para ello, en lugar de fijar una escala de análisis y una serie fija de cuadrantes de análisis, se tiene una serie aleatoria de zonas de análisis, las cuales se estudian a distintas escalas (con distintos tamaños). Para un proceso puntual dado, se trata de obtener una función que indique cuál es el numero de ocurrencias que deben darse a una distancia menor que un umbral dado h de cualquier punto generado por dicho proceso. La función que cumple esta definición se denomina función K [124], y puede expresarse como

$$K(h) = \frac{1}{\lambda}E(n) \qquad (13.4.9)$$

donde n es el número de eventos a distancia menor que h de un evento aleatorio cualquiera. La intensidad λ se añade para eliminar la influencia de la densidad, ya que el valor esperado de puntos a una distancia dada está en relación directa con dicha densidad.

Tiene sentido estudiar esta función tan solo para valores de h pequeños en comparación con el tamaño de la zona de estudio, ya que para otros valores no resulta coherente analizar los efectos de segundo orden dentro de dicha zona. Por ello, lo habitual es aplicar esta función solo a los valores de h menores que la mitad de la dimensión menor de la zona de estudio.

Un estimador de la función K es

$$\hat{K}(h) = \frac{1}{\lambda^2 A}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} I_h(d_{ij}) \qquad (13.4.10)$$

siendo I_h una función indicadora de la forma

$$I_h(d_{ij} = \begin{cases} 1 & \text{si } d_{ij} \leq h \\ 0 & \text{si } d_{ij} > h \end{cases} \qquad (13.4.11)$$

En este estimador no se consideran los efectos de borde, y aquellos puntos situados cerca de la frontera de la zona de estudio tendrán estimaciones inferiores a las reales. Un estimador que corrige estos efectos [124] es el siguiente:

$$\hat{K}(h) = \frac{1}{\lambda^2 A}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} \frac{I_h(d_{ij})}{w_{ij}} \qquad (13.4.12)$$

El valor w_ij pondera los distintos puntos en función de su distancia al borde de la zona de estudio. Para calcularlo se traza una circunferencia por el punto i con radio d_{ij} (es decir, una circunferencia con centro en el punto i y que pasa por el punto j), siendo w_{ij} la fracción de dicha circunferencia que queda dentro de la zona de estudio (Figura 13.7).

Hay que tener en cuenta que en ocasiones no es conveniente aplicar el efecto de borde, por ejemplo en el caso en que el proceso puntual subyacente no tenga lugar fuera de la zona de estudio.

Puesto que la densidad se estima como $\lambda = \frac{N}{A}$, la expresión del estimador de la función K queda finalmente como

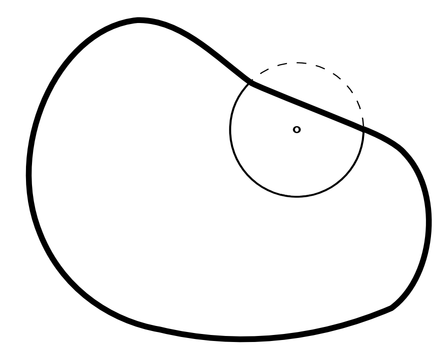


Figura 13.7: Corrección del estimador $\hat{K}(h)$ en función de los efectos de borde. El parámetro de corrección es el cociente entre la longitud interior (en trazo continuo) y la total de la circunferencia

$$\hat{K}(h) = \frac{A}{N^2}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\frac{I_h(d_{ij})}{w_{ij}} \qquad (13.4.13)$$

Para interpretar el significado de la función K, se tiene que, en condiciones de aleatoriedad espacial completa, el número de eventos a una distancia menor que h es πh^2. Esto es, $K(h) = \pi h^2$. Comparando los valores esperados con los estimados, se tiene que si $\hat{K}(h) < K(h)$ existe agrupamiento, mientras que si $\hat{K}(h) > K(h)$ existe regularidad en la distribución.

Para esta interpretación resulta más habitual utilizar un estimador $\hat{L}(h)$ de la forma

$$\hat{L}(h) = \sqrt{\frac{\hat{K}(h)}{\pi}} - h \qquad (13.4.14)$$

de tal modo que valores positivos de la misma indican agregación, mientras que los negativos indican regularidad.

Además de comparar el valor estimado con el valor esperado de la función K en condiciones de aleatoriedad espacial completa, puede compararse con el esperado para un proceso puntual determinado. Los valores de la función K son conocidos para muchos procesos puntuales, y esa información puede utilizarse para establecer comparaciones de igual modo. Distribuciones como las de Cox[125] o Gibbs han sido empleadas frecuentemente para el análisis de fenómenos tales como las distribuciones de pies dentro de masas forestales.

Frente a este enfoque, existe también la posibilidad de realizar un número n (preferiblemente grande) de simulaciones de un proceso y calcular la media y desviación típica de los valores de la función K obtenidos en ellas. Con ellos puede posteriormente calcularse la probabilidad de que una distribución observada de puntos represente un resultado generado por dicho proceso.

Al igual que los métodos restantes, el empleo de funciones K se realiza con carácter global, asumiendo la estacionaridad de la función $K(h)$. No obstante, puede adaptarse a un uso local, considerando en lugar de una serie de puntos aleatorios, un punto concreto i. La expresión 13.4.12 puede particularizarse para dar un estimador de esta función K local, según

$$\hat{K}(h) = \frac{1}{\lambda^2 A}\sum_{j=1,j\neq i}^{N}\frac{I_h(d_{ij})}{w_{ij}} \qquad (13.4.15)$$

13.5. Autocorrelación espacial

Como vimos en 11.2.3, la autocorrelación espacial indica la relación entre el valor de una variable existente en un punto dado y los de la misma variable en el entorno cercano de dicho punto. La autocorrelación espacial es la expresión formal de la primera ley geográfica de Tobler, y puede ser tanto positiva (los puntos cercanos exhiben valores más similares que los puntos lejanos) o negativa (los puntos lejanos exhiben valores más similares que los puntos cercanos).

El desarrollo realizado entonces se centraba en tratar las implicaciones que la existencia de autocorrelación espacial tiene para el análisis estadístico de datos espaciales. En este apartado veremos índices que permiten evaluar el grado de autocorrelación espacial existente, así como elementos mediante los cuales dicha autocorrelación podrá utilizarse posteriormente como parte integrante de otras formulaciones, en particular las relacionadas con interpolación (Capítulo 14).

13.5.1. La matriz de ponderación espacial

El concepto de autocorrelación espacial implica la definición de una *vecindad* de los distintos elementos geográficos. Se tiene que los valores de una variable registrados en aquellos elementos vecinos ejercen una influencia sobre los valores de dicha variable en un punto dado. Por ello es importante definir cuándo dos elementos son vecinos o no.

Aunque trabajamos con datos puntuales, este concepto de vecindad puede asociarse a otro tipo de entidades, como por ejemplo las de área. Así, puede considerarse que dos polígonos son vecinos si comparten al menos un lado común o, más restrictivamente, si comparten una longitud de sus perímetros mayor que un determinado umbral.

Para el caso de puntos, esta vecindad puede establecerse por distancia, considerando vecinos a todos aquellos puntos a una distancia menor que un umbral establecido. Este umbral puede aplicarse en todas direcciones (isotropía) o ser variable en función de la dirección (anisotropía).

De forma general, pueden considerarse todos aquellos factores que hagan que una entidad ejerza influencia sobre otra, y en el grado en la que dicha influencia tenga lugar. Esto puede incluir la consideración de otras relaciones existentes, como por ejemplo movimientos migratorios de especies, que «enlazan» unas entidades con otras y causan la existencia de interacción entre ellas más allá de la propia existente por distancia o contigüidad [126].

En la función K de Ripley ya vimos en la ecuación 13.4.10 cómo el uso del indicador I definía ese concepto de vecindad «efectiva», ya que tomaba valor cero para los puntos a una distancia mayor que h, haciendo que dichos puntos no tuvieran efecto sobre el resultado final de la función. De forma similar, puede extenderse el concepto de este indicador para construir la denominada *matriz de ponderación espacial.*

Para un conjunto de N entidades se tiene una matriz W de dimensiones $N \times N$ en la que el elemento w_{ij} refleja la influencia de la entidad i sobre la j. Por convención, los valores w_{ii} son iguales a cero. En el caso más sencillo, la matriz es de tipo binario, conteniendo únicamente valores 1 (existe vecindad efectiva entre las entidades) o 0 (no existe vecindad), pero los valores pueden ser cualesquiera. En la práctica, es de hecho habitual dividir estos valores por la suma de todos los valores de la columna, de forma que estén acotados siempre entre 0 y 1.

Mas allá de los valores que pueda contener, una característica primordial de la matriz de ponderación espacial es el método con el que ha sido creada, ya que la forma en la que se establece

la vecindad entre los distintos elementos tiene influencia directa sobre dicha matriz. Esto, sin duda, afecta a las operaciones realizadas posteriormente sobre esta, por lo que la elección del método a emplear en su creación es altamente relevante.

13.5.2. Medidas de autocorrelación espacial

Dos son las medidas más habituales para cuantificar la autocorrelación espacial de una variable: el parámetro I de Moran [127] y el parámetro c de Geary [128]. Ambos hacen uso de la matriz de ponderación espacial antes descrita.

En el caso del parámetro I de Moran, su expresión es

$$I = \frac{N}{S_0}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{w_{ij}(x_i-\mu)(x_j-\mu)}{\sum_{i=1}^{N}(x_i-\mu)^2} \quad (13.5.1)$$

donde μ es la media de la variable x y S_0 es un factor de normalización igual a la suma de todos los elementos de la matriz.

$$S_0 = \sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij} \quad (13.5.2)$$

Si los valores de la matriz han sido normalizados dividiéndolos por la suma de las columnas, $S_0 = N$ y la expresión anterior se simplifica.

$$I^* = \sum_{i=1}^{N}\sum_{j=1}^{N}\frac{w_{ij}(x_i-\mu)(x_j-\mu)}{\sum_{i=1}^{N}(x_i-\mu)^2} \quad (13.5.3)$$

El valor esperado de I es:

$$E(I) = \frac{-1}{N-1} \quad (13.5.4)$$

Valores por debajo de este valor esperado indican autocorrelación negativa, mientras que los situados por encima reflejan autocorrelación positiva. Al igual que sucede para otros parámetros, los valores de la desviación típica del parámetro I son conocidos, lo que permite establecer intervalos de confianza para rechazar o aceptar la hipótesis nula de ausencia de autocorrelación espacial.

Estos valores de la desviación típica tienen distintas expresiones en función de bajo qué supuestos se determinen. Estos supuestos y las expresiones resultantes no se tratarán aquí, pero pueden consultarse, por ejemplo, en [129].

Respecto el parámetro c de Geary, su expresión es

$$c = \frac{N-1}{2S_0}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{w_{ij}(x_i-x_j)^2}{\sum_{i=1}^{N}(x_i-\mu)^2} \quad (13.5.5)$$

Mientras que el parámetro I da una caracterización más global, el parámetro c es más sensible a las variaciones locales a distancia reducida.

El valor esperado de c es 1. Valores menores de 1 indican autocorrelación espacial positiva, mientras que los superiores indican una autocorrelación negativa.

Ambos parámetros son parte de una familia de estadísticos denotadas como Γ, de la forma

$$\Gamma = \sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}b_{ij} \quad (13.5.6)$$

Con este esquema pueden expresarse otros indicadores tales como los denominados *índices de conteo conjunto* (*joint count*) [129] u otros más específicos.

Todos estos parámetros caracterizan la autocorrelación espacial para el conjunto completo de puntos, es decir, para todo el área de estudio. Junto a estos, existen otros parámetros que miden la autocorrelación espacial a nivel local.

[130] proponen dos nuevos parámetros $G_i(d)$ y $G_i^*(d)$ que cuantifican si un punto dado i se encuentra rodeado por agrupaciones de puntos con valores altos o bajos. En el caso de $G_i(d)$ no se tiene en cuenta el valor del punto i mientras que en el caso de $G_i^*(d)$ sí se emplea este.

De forma similar, [131] propone una versión local del parámetro I de Moran, denotándolo como *indicador local de asociación espacial* (*Local Indicator of Spatial Association*, LISA).

La forma de interpretar estos parámetros locales es similar a lo visto anteriormente, y las formulaciones concretas de cada uno pueden consultarse en las referencias correspondientes.

13.5.3. Variogramas

Los variogramas son elementos clave para definir la autocorrelación espacial y aprovechar el conocimiento de esta dentro de formulaciones como el kriging (ver 14.2.4). Los variogramas se fundamentan en el concepto de semivarianza.

La semivarianza es una medida de la autocorrelación espacial de una variable x entre dos puntos i, j, y viene expresada por

$$\gamma(x_i, x_j) = \frac{1}{2}(z_i - z_j)^2 \qquad (13.5.7)$$

El cuadrado de las varianzas se multiplica por $\frac{1}{2}$ debido a que $\gamma(x_i, x_j) = \gamma(x_j, x_i)$. De ahí el uso del prefijo *semi*.

Puesto que puede calcularse la distancia entre dichos puntos, pueden representarse los valores de γ frente a las distancias h. Se obtiene una nube de puntos (nube del variograma) como la mostrada en la figura 13.8.

Esta nube aporta en principio poca información, pero puede resumirse agrupando los pares de puntos por intervalos de distancia, y calculando la media de todas las semivarianzas en cada intervalo. De esta forma se tiene una función que relaciona la semivarianza y la distancia entre puntos, según

$$\gamma(h) = \frac{1}{2m-(h)} \sum_{i=1}^{m(h)} (x_i - x_j)^2 \qquad (13.5.8)$$

siendo $m(h)$ el número de puntos del conjunto separados entre sí por una distancia h.

En la práctica se establecen una serie de valores de distancia equiespaciados, cada uno de los cuales define un intervalo centrado en dicho valor. La función $m(h)$ representa el número de puntos en cada bloque. Es importante que este número de puntos en cada bloque sea significativo, especialmente para dar validez al posterior ajuste sobre estos valores medios, como más adelante veremos.

La función $\gamma(h)$ es lo que se conoce como *variograma experimental*[1].

La nube de puntos de la figura 13.8 se resume en el variograma de la figura 13.9.

La elección de un tamaño óptimo para los intervalos es importante para obtener un variograma fiable. Si en el variograma aparecen ondulaciones, esto puede ser señal de que existe un comportamiento cíclico de la variable, pero más probablemente de que la distancia del intervalo no ha sido bien escogida.

Como puede verse en la figura, la curva que los puntos del variograma experimental describen implícitamente da lugar a la definición de unos elementos básicos que lo caracterizan.

- *Rango*. El rango representa la máxima distancia a partir de la cual existe dependencia espacial.

[1]Por emplear semivarianzas, es habitual también el uso del término *semivariograma*, aunque en general este se simplifica y se entiende que *variograma* hace referencia al elemento derivado de las semivarianzas. Será así como se emplee dentro de este texto.

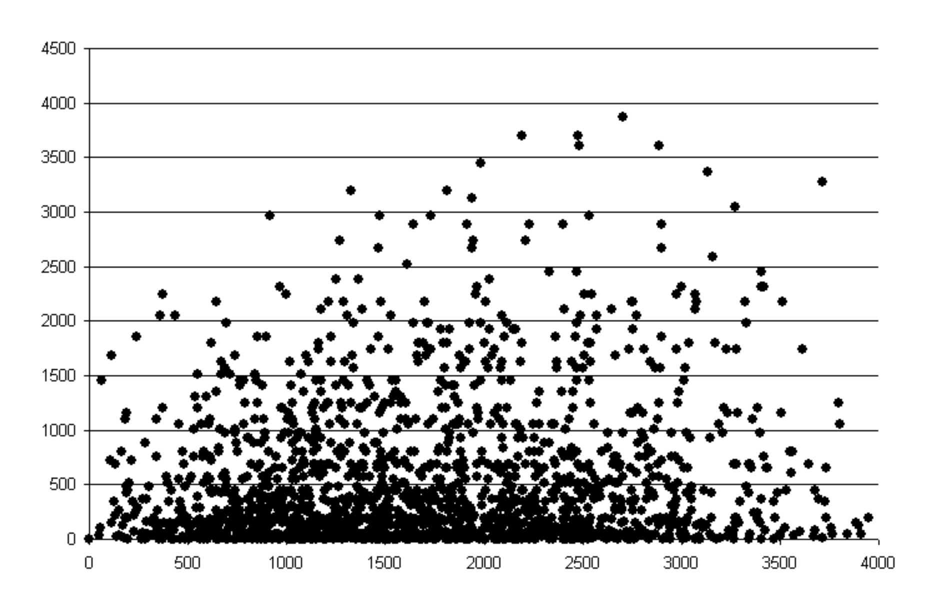


Figura 13.8: Representación de valores de semivarianza frente a distancia, formando la nube del variograma.

Es el valor en el que se alcanza la máxima varianza, o a partir del cual ya presenta una tendencia asintótica.

- *Sill*[2]. El máximo del variograma. Representa la máxima variabilidad en ausencia de dependencia espacial.
- *Nugget*. Conforme la distancia tiende a cero, el valor de la semivarianza tiende a este valor. Representa una variabilidad que no puede explicarse mediante la estructura espacial.

Por ejemplo, para el caso de la figura propuesta estos valores pueden estimarse aproximadamente a primera vista como rango $\simeq$ 3000, *sill* $\simeq$ 700 y *nugget* $\simeq$ 300.

Puesto que existen procesos para los cuales la variación de valores no se da igual en todas las direcciones, existen también *variogramas anisotrópicos* que no solo indican la variación media dentro de un intervalo de distancia, sino que caracterizan esa variación para una distancia y una dirección concreta.

Una forma de visualizar cómo la variación es distinta en función de la dirección considerada es a través de una *superficie variográficas* . Estas superficies no son mapas como tales (la superficie variográfica a partir de una capa ráster no tiene las mismas coordenadas que esta. De hecho, no tiene coordenadas absolutas en el espacio), sino que, respecto a un punto central en el cual la variación es lógicamente cero, expresan en cada celda el valor medio que se da a la distancia y dirección que dicha celda define respecto al punto central.

Si se traza un perfil de valores de esta superficie desde el punto central hasta un extremo de esta y en una dirección dada, el conjunto de dichos valores conforma el variograma particular de esa dirección.

[2]Tanto *Sill* como *Nugget* son términos ingleses que se emplean sin traducir de forma habitual, por lo que será así como se citen en este texto.

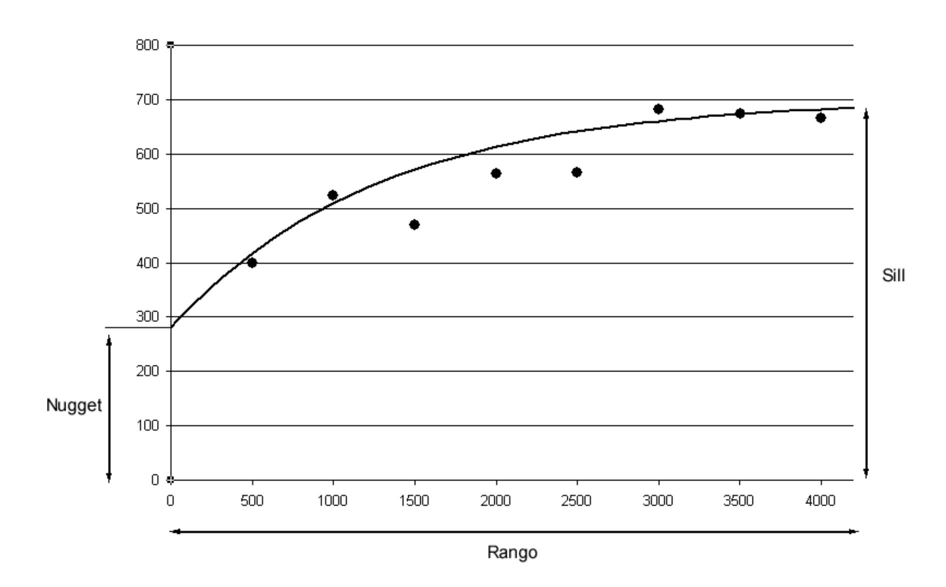


Figura 13.9: Resumen de la nube del variograma en un variograma experimental.

A partir de los puntos que forman el variograma experimental, puede definirse un modelo que aporta información sobre el proceso subyacente, a partir de su forma y sus parámetros. La definición de este modelo implica el ajuste de una curva a los puntos del variograma experimental, y tiene como resultado la obtención de un *variograma teórico*. En la figura 13.9 puede verse junto a los puntos del variograma experimental una curva ajustada a estos que define el variograma teórico. Sobre este último se pueden conocer las semivarianzas para cualquier distancia h, no solo para las definidas por los intervalos como en el caso del variograma experimental.

Existen muchas alternativas para elegir una función para el variograma teórico. Una función apta para este propósito debe cumplir los siguientes requisitos:

- Ser monótona creciente.
- Tener un máximo constante o asintótico. Es decir, un valor definido del *sill*. Funciones no acotadas superiormente, tales como las exponenciales, indicarían que la zona de estudio no es suficientemente grande, ya que no alcanza la dimensión a partir de la cual el efecto de la dependencia espacial deja de existir.
- $\gamma(0) > 0$. Es decir, el *nugget* debe ser positivo.

Sobre la base anterior, se pueden establecer familias principales de funciones aptas para definir un variograma teórico:

- *Nugget* puro. Funciones constantes de la forma

$$\gamma(h) = C_0 \qquad ; \qquad \forall h > 0 \qquad (13.5.9)$$

- Funciones que alcanzan el valor del *sill* (c) para un rango concreto (a). Son funciones de la forma

$$\gamma(h) = \begin{cases} f(x) & \text{si } h \leq a \\ c & \text{si } h > a \end{cases} \qquad (13.5.10)$$

siendo las más habituales de las funciones $f(x)$ las de tipo lineal, circular o esférico. Las formulas detalladas de estas y otras funciones pueden consultarse, por ejemplo en [132].

- Funciones que tienden asintóticamente al valor del *sill* (c). Se define un *rango efectivo*, siendo

este en el cual se da $\gamma(h) = 0{,}95c$. Entre estas funciones destacan

Exponencial

$$\gamma(h) = c\left(1 - e^{\frac{-3h}{a}}\right) \quad (13.5.11)$$

Gaussiana

$$\gamma(h) = c\left(1 - e^{\frac{-9h^2}{a^2}}\right) \quad (13.5.12)$$

Algunos autores como [133] usan h directamente en lugar de $3h$, en cuyo caso el rango no es igual a a, sino a $\frac{a}{3}$

El empleo de uno u otro modelo dependerá del conocimiento que tengamos acerca del proceso modelizado.

La figura 13.10 muestra las gráficas de los tipos de variogramas anteriores.

Llevar a cabo el ajuste del variograma teórico no es en absoluto un proceso trivial. Lo más sencillo es tratar de minimizar el error cuadrático. No obstante, deben tenerse en cuenta algunas consideraciones adicionales como las siguientes:

- No todos los puntos del variograma experimental son igual de precisos. Si en un intervalo solo había cinco puntos en la nube del variograma mientras que en otro había 50, debe favorecerse un ajuste correcto sobre este último antes que sobre el primero, ya que su precisión será mayor.
- Los puntos para valores altos del espaciamiento h son menos relevantes y debe darse más importancia en el ajuste a los relativos a valores bajos. Esto se debe a que el objeto del variograma es modelizar la influencia que ejercen los puntos cercanos, y más allá del valor del rango esa influencia no se da, con lo que no es una parte de interés del variograma. Asimismo, la aplicación del kriging se realiza utilizando la parte inicial del variograma (valores pequeños de h), especialmente cuando hay una gran densidad de datos, por lo que resulta más apropiado tratar de minimizar los errores en esta parte inicial.

Una solución para incorporar lo anterior es, en lugar de minimizar el error cuadrático total, minimizar este ponderado según el número de puntos en cada intervalo y las distancias de estos. Es decir, minimizar

$$\sum_{i=1}^{b} \frac{N_i}{h_i}(\hat{\gamma}(h_i) - \gamma(h_i))^2 \quad (13.5.13)$$

siendo b el número de intervalos, $\hat{\gamma}(h_i)$ el valor en el variograma experimental y $\gamma(h_i)$ el valor en el variograma teórico.

La inspección visual del ajuste es también importante y resulta conveniente llevarla a cabo.

Por último, es importante señalar que el número total de puntos considerados debe tenerse en cuenta para saber si el variograma teórico calculado es fiable o no. Aunque resulta imposible establecer fórmulas exactas al respecto, se acepta generalmente que con menos de 50 puntos la fiabilidad del variograma será dudosa. Valores entre 100 y 150 son adecuados, y mayores de 250 puntos garantizan un variograma fiable.

En el caso de tratarse de variogramas anisotrópicos, estos números son mayores.

13.5.4. Correlogramas

Para dos variables independientes x e y dadas, se define la covarianza de una muestra como

$$S_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) \quad (13.5.14)$$

Puede aplicarse este concepto para una única variable dada. Para dos puntos dados, su covarianza es

$$S_{ij} = (x_i - \overline{x})(x_j - \overline{y}) \quad (13.5.15)$$

El conjunto de valores de covarianza y distancias entre puntos da lugar a una nube de valores

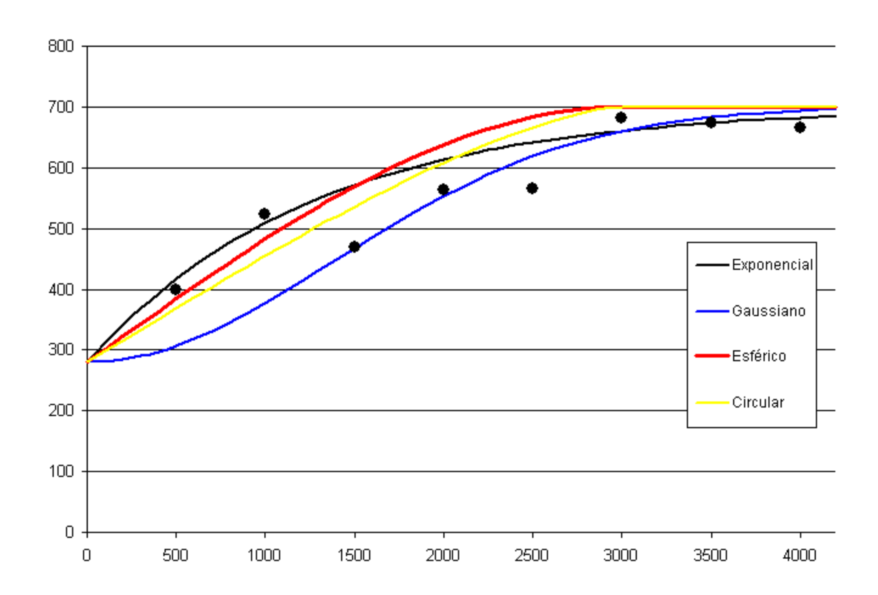


Figura 13.10: Distintos modelos de variograma teórico con los mismos parámetros de forma.

que, al igual que ocurría con las semivarianzas, puede emplearse para crear una curva experimental y a partir de esta una curva teórica. Con dicha curva teórica se tiene conocimiento de la covarianza a cualquier distancia, y recibe el nombre de *correlograma*.

Existe una relación directa entre el variograma y el correlagrama, como puede verse en la figura 13.11. Con la notación de la figura, se tiene para el caso del variograma que

$$\gamma(h) = \begin{cases} 0 & \text{si } \|h\| = 0 \\ C_0 + C_1\left(1 - e^{h/a}\right) & \text{si } \|h\| > a \end{cases}$$

Para el correlograma, se tiene que

$$\gamma(h) = \begin{cases} C_0 + C_1 & \text{si } \|h\| = 0 \\ C_1\left(e^{h/a}\right) & \text{si } \|h\| > a \end{cases} \qquad (13.5.16)$$

En la práctica, se emplea el variograma porque resulta más sencillo modelizar las semivarianzas que las covarianzas.

13.6. Resumen

Los datos espaciales presentan particularidades que deben considerarse a la hora de realizar cálculos estadísticos sobre ellos.Teniendo esto en cuenta, existen muy diversas formas llevar a cabo el análisis estadístico de datos espaciales, de las cuales hemos visto algunas de las más importantes.

Los elementos básicos de estadística descriptiva para datos espaciales son el centro medio, el centro mediano y la distancia típica. La elipse de variación permite representar gráficamente la dispersión, considerando que esta no se da igual en todas direcciones.

En el caso de trabajar con líneas y las direcciones que estas definen, es importante tener en cuenta la naturaleza circular de las variables. El trabajo con vectores en lugar de valores escalares es una solución práctica habitual para evitar resultados incorrectos.

Otro elemento importante del análisis estadístico espacial es el análisis de patrones de puntos.

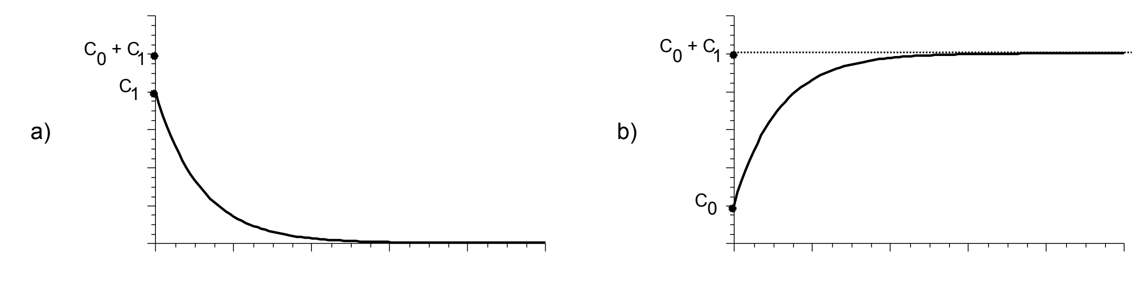


Figura 13.11: Relación entre correlograma (a) y variograma (b)

El método de división por cuadrantes, el de vecino más cercano, o el basado en funciones K de Ripley, todos ellos permiten caracterizar la disposición espacial de los puntos y con ello el proceso puntual inherente que da lugar a la misma.

Por último, la existencia de autocorrelación espacial puede medirse con índices como el I de Moran o el c de Geary, así como analizarse a través de variogramas. A partir de los datos de las semivarianzas se elabora un variograma experimental, el cual sirve como base para el ajuste de un variograma teórico. Este puede puede emplearse posteriormente en otras técnicas tales como el kriging, que veremos más adelante.

Capítulo 14

Creación de capas ráster

14.1. Introducción

Una buena parte de los análisis geográficos se realizan sobre capas ráster. Estas, por sus propias características, se prestan mejor a cierto tipo de análisis, y la implementación de estos resulta más sencilla, por lo que lo habitual es encontrarlos implementados para ser empleados con capas de esta clase.

La información de la que disponemos no siempre se encuentra en este formato, pero ello no significa que no podamos utilizarla. A partir de información almacenada según otros modelos de datos, podemos generar capas ráster (este proceso lo denominaremos *rasterización*) que reúnan las características necesarias para usarse con los algoritmos correspondientes. El problema es en todos los casos la creación de una estructura regular a partir de información que no es regular, tal como la contenida en un TIN, una capa de polígonos, una de líneas, o una capa de valores puntuales.

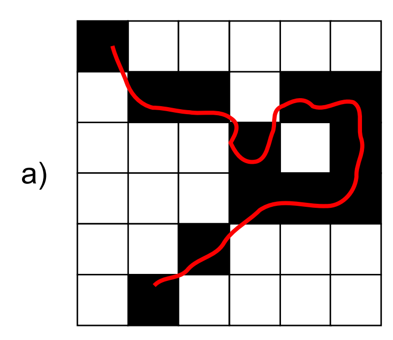
Si disponemos de una capa de polígonos y estos cubren la totalidad del territorio, este proceso no es difícil. Basta ver dentro de qué polígono cae la coordenada que define cada celda, y asignar a esta el valor de uno de los atributos de la capa de polígonos, el cual contenga la variable a recoger en la capa ráster. En el caso del TIN es similar, ya que cada uno de los triángulos permite el cálculo de valores en sus puntos, y puede igualmente establecerse una relación entre estos y las celdas de una malla ráster que cubra el mismo espacio geográfico.

Si tenemos una capa de líneas la cosa no es muy distinta. Basta ver por qué celdas pasan esas líneas y asignar el valor de estas a dichas celdas. Pueden existir ambigüedades a la hora de considerar *cuánto* ha de recorrer una línea a través de una celda para considerar que pasa por esta y asignarle el valor correspondientes, como se muestra en la figura 14.1. Aun así, y salvando estos aspectos, no resulta difícil rasterizar una capa de líneas y tener una capa ráster válida.

Por último, para el caso de una capa de puntos, la conversión en una capa ráster es aún más sencilla. Basta con asignar los correspondientes valores a aquellas celdas en las que se sitúen los puntos. Las restantes, y puesto que no existe información al respecto, deberán llevar el valor que codifica la ausencia de datos.

Estas formas de rasterización, sin embargo, pueden no ser idóneas según las circunstancias. En los dos últimos casos (líneas y puntos), y especialmente para el caso de puntos, la situación que se recoge en la capa ráster puede no ser la necesaria para el análisis, y realmente no estaremos aprovechando las capacidades del modelo de datos ráster a pesar de haber llevado a cabo una conversión.

Por ejemplo, si las líneas que rasterizamos son curvas de nivel, van a indicar la elevación en



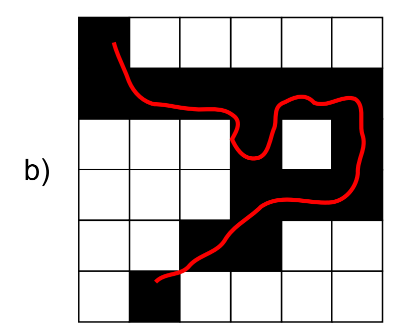


Figura 14.1: Formas distintas de rasterizar una línea en función del criterio empleado.

las mismas. Fuera de ellas, la capa ráster generada no tendrá datos, pero en realidad esas celdas sí que tienen una elevación concreta. Del mismo modo sucede si medimos esa elevación en una serie de puntos en lugar de en líneas y después rasterizamos la capa, o si medimos cualquier otra variable en un número dado de localizaciones puntuales y hacemos lo propio.

Para aprovechar la buena disposición del modelo ráster para el análisis, y especialmente en el caso de variables continuas, debemos tener una capa que contenga información en todas sus celdas, incluso si originalmente solo hemos medido los valores de la variable estudiada en una serie de celdas. Será en base a los valores de esas celdas como obtengamos posteriormente los valores en las restantes.

Los métodos de interpolación que veremos en este capítulo permiten rellenar esas celdas restantes a partir de los valores puntuales conocidos, realizando estimaciones. Para ello, aplican conceptos de estadística espacial más o menos complejos según su formulación particular, de tal modo que los puntos cercanos a cada celda son los que determinan el valor estimado de esta. Este hecho es una aplicación directa de la ley de Tobler, que establece que los puntos cercanos tiene mayor probabilidad de tener valores similares que aquellos separados por una distancia mayor.

En términos generales, un método de interpolación es una herramienta que permite el cálculo del valor de una variable en una coordenada para la cual dicho valor no es conocido, a partir de los valores conocidos para otra serie de coordenadas. En el caso particular de la creación de una capa ráster, las coordenadas (x, y) donde han de calcularse los valores desconocidos son los centros de las celdas sin dato recogido.

El número de métodos distintos es muy amplio, y es importante reseñar que la bondad o no de uno u otro va ligada no solo al método en sí, sino también a la variable interpolada y al uso que se dé posteriormente a la capa resultante.

La aplicación de los métodos de interpolación es, asimismo, muy diversa. Casos habituales son, por ejemplo, la realizada a partir de datos de elevación tomados en campo mediante GPS o estación total, o de los datos climatológicos de precipitación y temperatura registrados en los observatorios de una red. Resulta imposible recoger valores para cada una de las celdas de una capa ráster que cubra el territorio a estudiar, pero estas variables, por ser continuas, se manejarán mejor y serán más útiles si se dispone de ellas en formato ráster. Los métodos de interpolación son los encargados de convertir esos datos puntuales en mallas regulares.

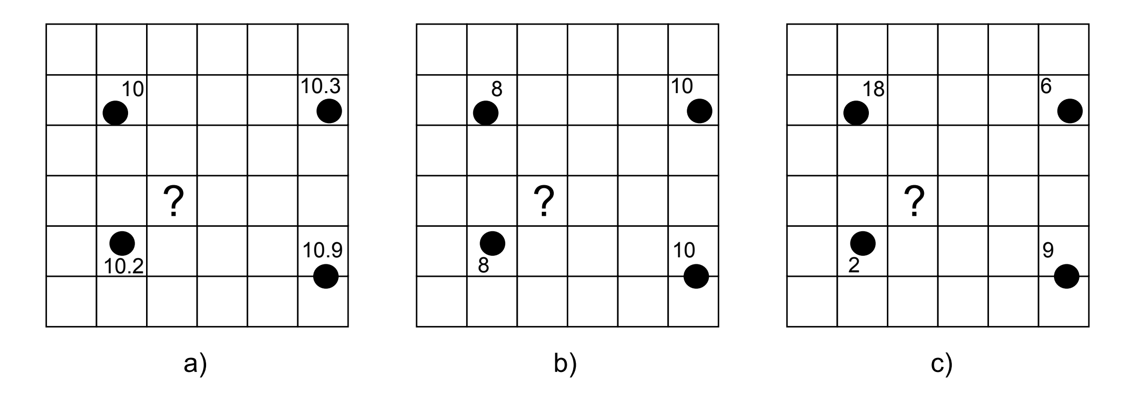


Figura 14.2: Tres situaciones para la aplicación de interpolación espacial.

En general, cualquier variable recogida mediante muestreo puede ser el punto de partida para la aplicación de dichos métodos.

En algunos de los casos anteriores, los valores en las celdas guardan una dependencia no solo con los puntos que contienen la variable interpolada, sino también con otras variables de las que puede o no disponerse de información. Por ejemplo, para el caso de la temperatura, esta se encuentra influenciada por la elevación. De igual modo, las características del suelo tendrán relación con parámetros del relieve tales como la pendiente o índices relacionados con la humedad topográfica (Capítulo 16).

Estas variables, que denominamos *variables de apoyo* o *predictores*, puede incorporarse como datos de partida a algunos métodos de interpolación, aumentando así su precisión. En general, se requiere que dichas variables de apoyo estén recogidas en formato ráster. Asimismo, pueden plantearse análisis de regresión mediante los cuales, y sin necesidad de utilizar la componente espacial, puedan estimarse los valores en las celdas problema a partir de los valores de la variable en los puntos conocidos y los valores de los predictores tanto en dichos puntos como en las celdas a rellenar.

Junto con lo anterior, la información de una determinada variable cuantitativa tomada en ciertos puntos puede servir para estimar densidades de dicha variable (tales como, por ejemplo, individuos de una especie) y crear superficies continuas. Este análisis se lleva a cabo no con métodos de interpolación o regresión, sino con otra serie de algoritmos habituales en los SIG que veremos al final del capítulo.

14.2. Interpolación

Un método de interpolación permite el calculo de valores en puntos no muestreados, a partir de los valores recogidos en otra serie de puntos.

En la figura 14.2 encontramos tres situaciones en las cuales se hace necesario el uso de interpolación si queremos obtener una capa raster con valores en todas sus celdas. En dichas situaciones, los cuatro puntos señalados han sido muestreados y se dispone de un valor en ellos. Adviértase que no han de encontrarse necesariamente en el centro de las celdas. El problema es el mismo en cada uno de estos supuestos: queremos estimar los valores en las celdas de la malla, y en particular en la celda marcada con un interrogante.

Para el caso a) la lógica nos indica que el valor en esta celda debe estar alrededor de 10,

ya que este valor sigue la tendencia natural de los valores recogidos, que tiene todos ellos un valor de esa magnitud. Si aplicamos cualquiera de los métodos de interpolación que veremos a continuación, el valor que obtengamos será con seguridad muy aproximado a esa cifra.

Supongamos ahora que nos encontramos en el caso b). En este supuesto, la lógica nos indica que el valor será inferior a 10, y también probablemente a la media de los valores muestrales (9), ya que la celda problema se sitúa más cerca de los valores inferiores que de los superiores a ese valor medio. Razonando de este modo, aplicamos el hecho de que la proximidad incrementa la semejanza de valores. Es decir, que existe autocorrelación espacial para la variable interpolada.

El caso c) es algo distinto. Aquí ya no parece tan sencillo «adivinar» el valor que corresponde. Esto es así no porque las operaciones sean más complejas, sino porque no existe de la misma forma que en los ejemplos anteriores la autocorrelación espacial de la variable, y esa *lógica* no resulta tan obvia. Utilizando los distintos métodos de interpolación, puede ser que estos den valores distintos, ya que se comportarán de forma diferente ante tal situación.

Estos sencillos ejemplos numéricos tienen como objetivo dar imagen general de lo que el proceso de interpolación conlleva, y que puede resultar más intuitivo al analizarlo sobre un conjunto reducido de puntos. A pesar de sus diferencias, grandes en muchos casos, todos parten de ideas comunes que lo único que pretenden es replicar de forma lo más precisa posible un campo a partir de un conjunto definido de puntos con valores de este.

Existen muchos métodos de interpolación, de los cuales algunos cuentan con más presencia en los SIG por estar más adaptados al tipo de dato que se maneja. Su aplicación habitual dentro de un SIG es bidimensional, ya que una capa ráster es una entidad bidimensional. Hablamos, por ello, de *interpolación espacial*. No obstante, estos métodos no han de restringirse al plano, y pueden extenderse a un numero superior de dimensiones para reflejar otras variables tales como la profundidad (por ejemplo, para construir un modelo tridimensional de las características del suelo entre dos profundidades establecidas y con un intervalo dado), o bien el tiempo.

Podemos clasificar los distintos métodos de interpolación según varios criterios [134].

- **Según los puntos considerados para el cálculo de valores**. Algunos métodos consideran que todos los puntos de los que disponemos tienen influencia sobre el valor a calcular en una celda. Estos modelos se conocen como *globales*. En otros, denominados *locales*, solo se considera un conjunto restringido de estos. Este conjunto puede establecerse por medio de un umbral de distancia (todos los situados a una distancia menor que el umbral), de conteo (los n puntos más cercanos), o bien ambos.
  La selección de este conjunto de puntos más cercanos (los de más influencia) es un aspecto importante en el rendimiento de los métodos de interpolación de este tipo. Cuando se trabaja con un número de puntos elevado, se hace inviable el cálculo de las distancias entre todos esos puntos para seleccionar los más cercanos. El uso de índices espaciales y otras estructuras semejantes (véase 12.4.2) se hace necesario para poder aplicar eficientemente estos métodos de interpolación sobre dichos conjuntos con tal número de puntos.
  En realidad, un método global puede entenderse como uno local con un umbral infinito, no existiendo una dicotomía estricta entre ambas clases.
- **Según su valor en los puntos de partida**. En algunos métodos, denominados *exactos*, los valores asignados a las coordenadas correspon-

dientes a los puntos de origen son exactamente los recogidos en dichos puntos. En los métodos *aproximados*, el valor en esas celdas es el que corresponde al mejor ajuste, y no ha de coincidir necesariamente con el valor original.

- **Según la inclusión o no de elementos probabilísticos**. Diferenciamos entre métodos *estocásticos* (aquellos que emplean elementos probabilísticos) y métodos *determinísticos* (aquellos que no los emplean).

14.2.1. Por vecindad

El método más sencillo de interpolación es el de *vecindad* o *vecino más cercano*. En él se asigna directamente a cada celda el valor del punto más cercano. No existe formulación matemática que emplee las distancias entre puntos o los valores de estos, sino que el valor resultante es sencillamente el del punto más próximo.

Se trata, por tanto, de un método local, exacto y determinístico.

El resultado es una capa con saltos abruptos (tanto como lo sean las diferencias entre los valores de puntos cercanos), con un aspecto «aterrazado» (Figura 14.3). El conjunto de celdas con el mismo valor (dentro de la misma terraza) representa el lugar geométrico de las celdas cuyo punto más cercano de entre los de partida es uno dado.

La interpolación por vecindad no es adecuada para el trabajo con variables continuas, pero sí para variables categóricas. Por ejemplo, para un conjunto de puntos cada uno de los cuales esté identificado con un código numérico, la interpolación por vecindad de ese valor da como resultado una capa donde los valores de las celdas identifican el punto más cercano. Esto puede utilizarse para calcular la influencia de cada uno de ellos en el espacio representado.

Este tipo de razonamientos ha sido empleado tradicionalmente para calcular los denominados *polígonos de Thiessen*, de uso habitual en el análisis climatológico, asociando a cada zona los valores de la estación meteorológica más cercana. Estos polígonos de Thiessen conforman una estructura conocida como *teselación de Voronoi*, que puede también calcularse de forma vectorial, como veremos en el capítulo 18. La teselación de Voronoi está íntimamente ligada a la denominada *triangulación de Delaunay*, base para la construcción de TIN como en su momento detallaremos.

14.2.2. Métodos basados en ponderación por distancia

Los métodos basados en ponderación por distancia son algoritmos de interpolación de tipo local, aproximados y determinísticos. El valor en una coordenada dada se calcula mediante una media ponderada de los puntos de influencia seleccionados (bien sea la selección por distancia o por número de estos). Su expresión es de la forma [135]

$$\widehat{z} = \frac{\sum_{i=1}^{n} p_i z_i^k}{\sum_{i=1}^{n} p_i^k} \qquad (14.2.1)$$

siendo p_i el peso asignado al punto i-ésimo. Este peso puede ser cualquier función dependiente de la distancia.

La función más habitual es la que da lugar al método de *ponderación por distancia inversa*, de la forma

$$p_i = \frac{1}{d_i^k} \qquad (14.2.2)$$

donde el exponente k toma habitualmente el valor 2.

Otra función habitual es la de la forma

$$p_i = e^{-kd_i} \qquad (14.2.3)$$

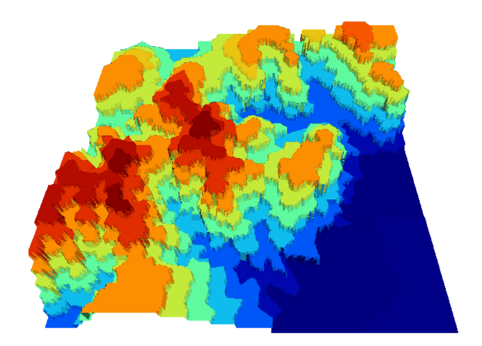


Figura 14.3: Superficie obtenida mediante interpolación por vecindad.

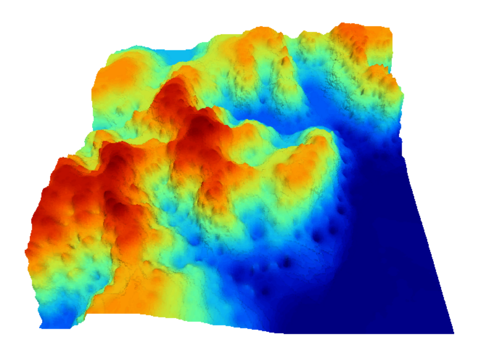


Figura 14.4: Superficie obtenida mediante ponderación por distancia inversa

así como la que constituye el método de *decremento lineal*, con la siguiente expresión:

$$p_i = 1 - (\frac{d_i}{d_{max}})^k \tag{14.2.4}$$

donde k es un parámetro que regula la forma de la función y d_{max} la distancia máxima de influencia.

La figura 14.4 muestra la superficie calculada a partir de datos puntuales de elevación aplicando el método de ponderación por distancia inversa.

Los métodos basados en ponderación por distancia solo tienen en cuenta el alejamiento, pero no la posición. Es decir, un punto situado a una distancia d hacia el Norte tiene la misma influencia que uno situado a esa misma distancia d pero hacia el Oeste.

Igualmente, los métodos basados en distancia no generan valores que se encuentren fuera del rango de valores de los datos de entrada. Eso causa efectos indeseados en caso de que el muestreo de datos no recoja los puntos característicos de la superficie interpolada. La figura 14.5 muestra en un caso unidimensional cómo, de no recogerse los valores extremos (tales como cimas o valles), estos desaparecen y son sustituidos por tramos llanos

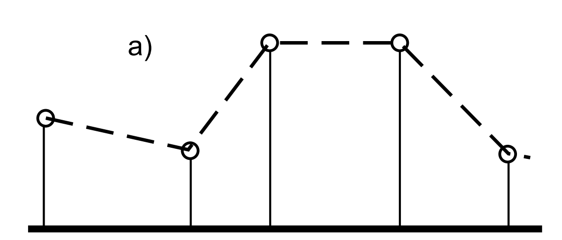
Puede entenderse el método de vecino más cercano como un caso particular de método ponderado por distancia, en el que se emplea un único punto de influencia, y su peso asignado es $p_1 = 1$.

14.2.3. Curvas adaptativas (Splines)

Las curvas adaptativas o *splines* conforman una familia de métodos de interpolación exactos, determinísticos y locales. Desde un punto de vista físico pueden asemejarse a situar una superficie elástica sobre el área a interpolar, fijando esta sobre los puntos conocidos. Crean así superficies suaves, cuyas características pueden regularse modificando el tipo de curva empleada o los parámetros de esta, de la misma forma que sucedería si se variasen las cualidades de esa membrana ficticia.

La superficie creada cumple la condición de minimizar con carácter global alguna propiedad tal como la curvatura.

Desde un punto de vista matemático, los splines son funciones polinómicas por tramos, de tal modo que en lugar de emplear un único polinomio para ajustar a todo un intervalo, se emplea uno distinto de ellos para cada tramo. Las curvas definidas por estos polinomios se enlazan entre



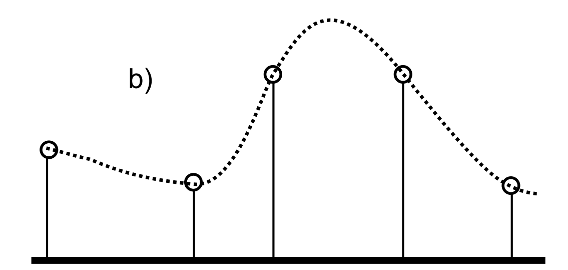


Figura 14.5: La interpolación basada en ponderación por distancia (a) no crea valores extremos si estos no han sido recogidos, lo cual deriva en un «aplanamiento» de la superficie y la aparición de falsas terrazas. Otros métodos tales como los *splines* (b) sí que permiten la aparición de valores fuera del rango muestreado.

ellas en los denominados *nudos*, cumpliendo unas condiciones particulares de continuidad.

Los splines no sufren los principales defectos de los dos métodos anteriores. Por un lado, pueden alcanzar valores fuera del rango definido por los puntos de partida. Por otro, el mal comportamiento de las funciones polinómicas entre puntos se evita incluso al utilizar polinomios de grados elevados. No obstante, en zonas con cambios bruscos de valores (como por ejemplo, dos puntos de entrada cercanos pero con valores muy diferentes), pueden presentarse oscilaciones artificiales significativas. Para solucionar este problema, una solución es el empleo de *splines con tensión* [136]. La incorporación de la tensión en una dirección permite añadir anisotropía al proceso de interpolación [137].

14.2.4. Kriging

El *kriging*[1] es un método de interpolación estocástico, exacto, aplicable tanto de forma global como local. Se trata de un método complejo con una fuerte carga (geo-)estadística, del que existen además diversas variantes.

El kriging se basa en la teoría de variables regionalizadas, la cual fue desarrollada por [138] a partir del trabajo pionero de [139]. El objetivo del método es ofrecer una forma objetiva de establecer la ponderación óptima entre los puntos en un interpolador local. Tal interpolación óptima debe cumplir los siguientes requisitos, que son cubiertos por el kriging:

- El error de predicción debe ser mínimo.
- Los puntos cercanos deben tener pesos mayores que los lejanos.
- La presencia de un punto cercano en una dirección dada debe restar influencia (enmascarar) a puntos en la misma dirección pero más lejanos.
- Puntos muy cercanos con valores muy similares deben «agruparse», de tal forma que no aparezca sesgo por sobremuestreo.
- La estimación del error debe hacerse en función de la estructura de los puntos, no de los valores.

Junto con la superficie interpolada, el kriging genera asimismo superficies con medidas del error de interpolación, que pueden emplearse para conocer la bondad de esta en las distintas zonas.

En su expresión fundamental, el kriging es semejante a un método basado en ponderación por distancia. Dicha expresión es de la forma

[1] *Krigeage* en su denominación original en francés, que se traduce como *krigeado* en castellano, aunque es mucho más habitual el uso de la denominación inglesa, la cual emplearemos aquí.

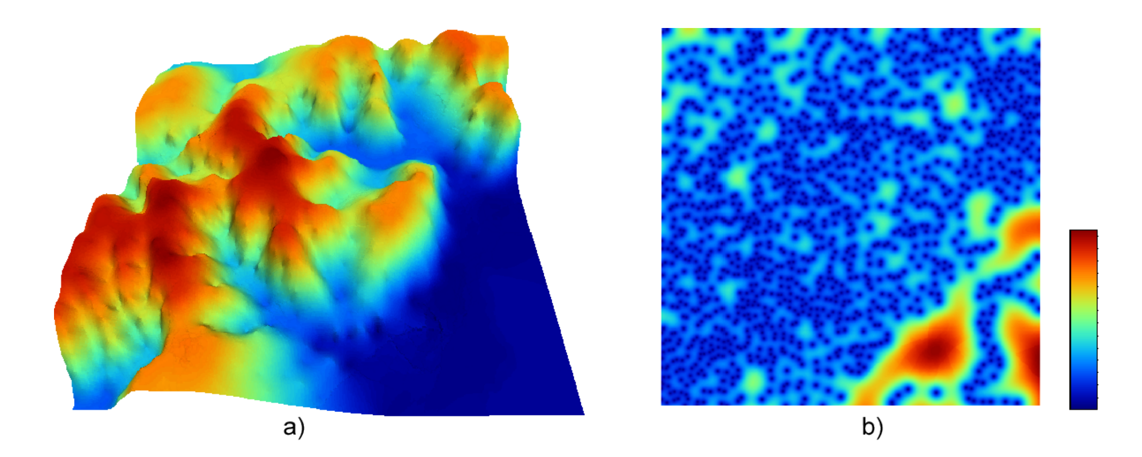


Figura 14.6: Superficie obtenida mediante interpolación por kriging ordinario y capa de varianzas.

$$\widehat{z} = \sum_{i=1}^{n} z_i \Lambda_i \tag{14.2.5}$$

siendo Λ_i los pesos asignados a cada uno de los puntos considerados. El cálculo de estos pesos, no obstante, se realiza de forma más compleja que en la ponderación por distancia, ya que en lugar de utilizar dichas distancias se acude al análisis de la autocorrelación a través del variograma teórico (13.5.3). Por ello se requiere, asimismo, que exista un número suficiente de puntos (mayor de 50, como se vio en el capítulo 13) para estimar correctamente el variograma.

A partir de los valores del variograma, se estima un vector de pesos que, multiplicado por el vector de valores de los puntos de influencia, da el valor estimado.

En el denominado *kriging ordinario*, e interpolando para un punto p empleando n puntos de influencia alrededor de este, el antedicho vector de pesos se calcula según

$$\begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \\ \lambda \end{pmatrix} = \begin{pmatrix} \gamma(d_{11}) & \cdots & \gamma(d_{1n}) & 1 \\ \gamma(d_{21}) & \cdots & \gamma(d_{2n}) & 1 \\ \vdots & \ddots & \vdots & 1 \\ \gamma(d_{n1}) & \cdots & \gamma(d_{nn}) & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} \gamma(d_{1p}) \\ \gamma(d_{2p}) \\ \vdots \\ \gamma(d_{np}) \\ 1 \end{pmatrix} \tag{14.2.6}$$

siendo d_{ij} la distancia entre los puntos i y j, y λ un multiplicador de Lagrange.

El error esperado en la estimación también se obtiene, como se dijo antes, a partir de la estructura de los puntos. Es decir, utilizando igualmente el variograma. Para ello se emplean los pesos calculados anteriormente, según

$$S_p^2 = \sum_{i=1}^{n} w_i \gamma(d_{ip}) + \lambda \tag{14.2.7}$$

La aplicación del kriging ordinario implica la asunción de una serie de características de los datos:

- Estacionaridad de primer y segundo orden. La media y la varianza son constantes a lo largo del área interpolada, y la covarianza depende únicamente de la distancia entre puntos.
- Normalidad de la variable interpolada.
- Existencia de una autocorrelación significativa.

La figura 14.6 muestra una superficie obtenida mediante kriging ordinario, junto a la capa de varianzas asociada.

Cuando no puede asumirse la estacionariedad de primer orden y existen una tendencia marcada en el valor medio esperado en los distintos puntos, puede aplicarse un tipo de kriging denominado *kriging universal*. Además de los valores a interpolar y sus coordenadas, este método permite el uso de predictores relacionados con dicha tendencia.

El *kriging con regresión* es similar en cuanto a sus resultados e ideas, aunque la forma de proceder es distinta. Mientras que en el universal se trabaja con los residuos y la superficie de tendencia conjuntamente, este separa ambas partes y las analiza por separado, combinándolas después para estimar los valores y los errores asociados.

Existen muchas otras variaciones del kriging tales como el *kriging simple*, el *kriging por bloques* o el *co–kriging*. La aplicación de estas formas, sin embargo, es limitada, debido a que no es tan frecuente su implementación. Los SIG habituales implementan por regla general las variantes básicas anteriores, quedando las restantes para programas especializados.

14.2.5. Ajuste de funciones. Superficies de tendencia

El ajuste de funciones es un método de interpolación determinístico o estocástico (según el tipo de función a ajustar), aproximado y global. Puede aplicarse de forma local, aunque esto resulta menos habitual. Dado el conjunto completo de los puntos de partida, se estima una superficie definida por una función de la forma

$$\widehat{z} = f(x, y) \tag{14.2.8}$$

El ajuste de la función se realiza por mínimos cuadrados.

Estas funciones son de tipo polinómico, y permiten el cálculo de parámetros en todas las celdas de la capa ráster. Por su propia construcción, requieren pocas operaciones y son rápidos de calcular. Sin embargo, esta sencillez es también su principal inconveniente. Los polinomios de grado cero (plano constante), uno (plano inclinado), dos (colina o depresión) o tres, son todos ellos demasiado simples, y las variables continuas que pueden requerir un proceso de interpolación dentro de un SIG son por lo general mucho más complejas. Pueden emplearse polinomios de mayor grado que aumentan la precisión del ajuste en los puntos de partida. Sin embargo, aumenta por igual la oscilación de la función entre puntos, mostrando un mal comportamiento con grados elevados, y no obteniéndose en ningún caso la fidelidad a la superficie real que se logra con otros métodos.

La figura 14.7 muestra superficies con valores de elevación obtenidos mediante ajuste polinómico de grados dos y cinco.

El empleo de funciones de ajuste permite incorporar otras variables adicionales h_i mediante funciones de la forma

$$\widehat{z} = f(x, y, h_1, \ldots, h_n) \tag{14.2.9}$$

Esto posibilita la incorporación de variables de apoyo (predictores) que pueden tener una influencia en la variable interpolada, considerando así no únicamente la posición de los distintos puntos, sino los valores en ellos de dichas variables de apoyo.

Si bien en muchos casos la superficie interpolada puede no reflejar con precisión una variable, puede emplearse para considerar el efecto de tendencias globales en la distribución de esta. Una variable puede estar condicionada por efectos globales y efectos locales. Puesto que el ajuste de una función polinómica tal y como lo hemos

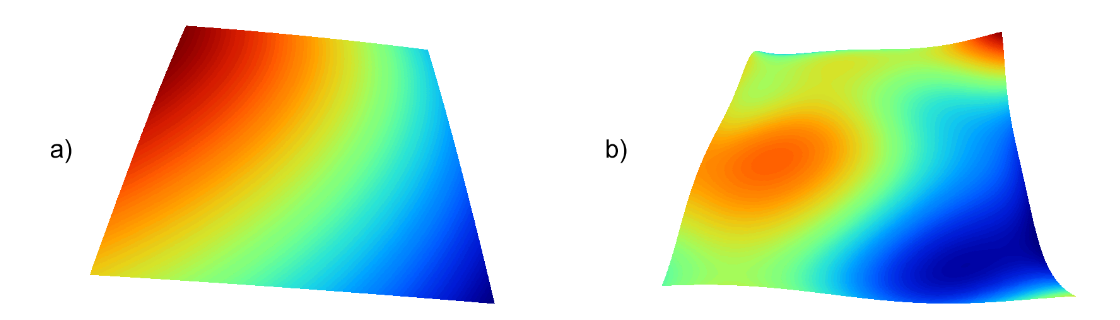


Figura 14.7: Superficies obtenidas mediante interpolación por ajuste polinómico de segundo (a) y quinto (b) grado.

visto aquí es un interpolador global, permite la separación de estos dos tipos de efectos. Este proceso constituye el empleo de las denominadas *superficies de tendencia*.

Supongamos una capa de temperatura. Esta dependerá de la altura, pero también de la latitud. Existe, por tanto, una tendencia global: a mayor latitud (supóngase hemisferio norte), menor temperatura. Esta se puede modelizar con una función lineal, y separar el efecto de ambos factores, altitud y latitud (Figura 14.8). El efecto local puede entenderse como el residuo de primer orden de la superficie interpolada.

En la aplicación de predictores debe tenerse en cuenta el principio de parsimonia: mantener el modelo lo más simple posible. La incorporación de nuevos predictores, si estos se encuentran significativamente correlacionados, conlleva un aumento de la multicolinearidad [140]. Esta circunstancia da lugar a un sobreajuste de la función y empeora la calidad de las estimaciones, especialmente en la predicción de datos fuera del área delimitada por los puntos de partida, es decir, la extrapolación.

Un caso particular de las funciones del tipo señalado en la ecuación 14.2.9 son las de la forma

$$\widehat{z} = f(h_1, \ldots, h_n) \qquad (14.2.10)$$

es decir, aquellas que no consideran la componente geográfica y simplemente realizan una regresión en el espacio de atributos.

Puesto que existe autocorrelación espacial, en el caso de considerar el espacio geográfico resulta más conveniente el uso de mínimos cuadrados generalizados en lugar de mínimos cuadrados ordinarios. En el capítulo 21 veremos con más detalle los temas relacionados con regresiones e inferencia estadística basada en datos espaciales.

14.2.6. Muestreo de datos para interpolación

Muchas veces, la información de la que disponemos la obtenemos de una fuente ajena y no queda más remedio que emplearla tal y como la recibimos. Tal es el caso de una curvas de nivel digitalizadas de un mapa clásico (no podemos disponer de más curvas porque no existen en el mapa original), o de unos datos de observatorios meteorológicos (no podemos densificar esa red y disponer de datos medidos en otros puntos).

En otras ocasiones, sin embargo, los datos pueden no haber sido tomados aún y debe plantearse un esquema para ello. Es decir, que ese conjunto finito de puntos a partir de los cuales obtener una superficie interpolada puede ser creado a voluntad, siempre que nos mantengamos dentro

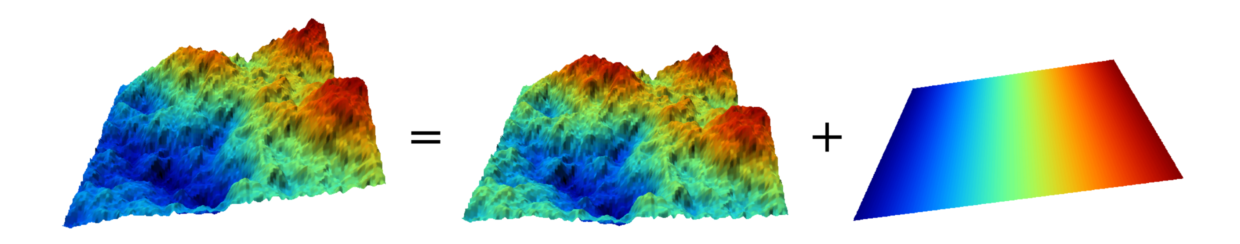


Figura 14.8: Separación de una variable en un efecto debido a una tendencia geográfica (en este caso en forma de plano inclinado con valores mayores en la parte derecha) y un efecto local.

del marco de limitaciones (temporales, monetarias, etc.) existentes. El diseño del muestreo de forma eficiente ayuda de manera directa a obtener mejores resultados al interpolar los datos recogidos. Este diseño de muestreo puede realizarse no solo para obtener un conjunto de datos base, sino para enriquecer uno ya existente, ampliándolo con nuevos puntos.

El objetivo de un muestreo espacial es poder inferir nueva información acerca de una población en la que sus distintos elementos están georreferenciados, a partir de un subconjunto de dicha población. La razón de llevar a cabo el muestreo es la imposibilidad de analizar todos los miembros de la población, ya que estos pueden ser muy numerosos (como en el caso de todos los habitantes de un país) o infinitos (como sucede para variables continuas tales como elevaciones o temperaturas).

Aunque el uso de un muestreo no es exclusivamente la interpolación posterior de sus datos para obtener una superficie continua, sí que es cierto que la interpolación se ha de basar en un muestreo previo, por lo que ambos conceptos van íntimamente unidos. Por ello, analizaremos en esta sección algunas de las ideas fundamentales que deben conocerse para diseñar un muestreo, y que directamente condicionan la calidad de la interpolación aplicada sobre el conjunto de puntos muestreados.

Lógicamente, el trabajar con una fracción reducida de la población introduce un cierto nivel de error. Ese error puede controlarse y acotarse mediante el diseño del muestreo. Por ello, a la hora de diseñar un muestreo deben plantearse principalmente las siguientes preguntas [141]:

- **¿Qué se pretende estimar?** Algunas propiedades no poseen componente espacial, y lo importante es el *cuánto*, no el *dónde*. Por ejemplo, la media de una variable. En el caso de la interpolación, sí es de gran importancia el dónde, puesto que, como ya sabemos, existe autocorrelación espacial y esta es utilizada como concepto inherente a los distintos métodos de interpolación.
- **¿Que tamaño de muestreo (número n de muestras) se necesita para un nivel de precisión dado?** El error está acotado en función de los puntos muestrales. En algunos casos se necesita realizar una estimación con un alto grado de precisión, y dentro de un amplio intervalo de confianza. En otros, sin embargo, no es necesario, y plantear un muestreo demasiado amplio supone un gasto innecesario de recursos.
- **¿Dónde tomar esas n muestras?** Por la existencia de autocorrelación espacial o por la variación espacial que pueden presentar otras variables relacionadas, una de las partes más impor-

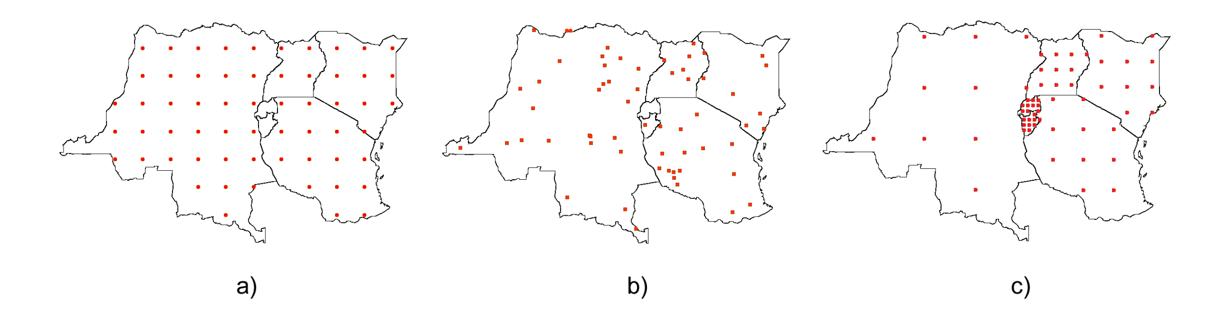


Figura 14.9: Tipos de muestreo. a) regular, b) aleatorio, c) estratificado

tantes del muestreo es determinar la localización exacta en la que tomar las muestras. Existen tres modelos básicos de muestreo: aleatorio, estratificado y regular (Figura 14.9)

El muestreo aleatorio simplemente sitúa los puntos de muestreo al azar sin obedecer a ningún condición particular. Si se desconoce el comportamiento de la variable muestreada, un muestreo aleatorio puede ser una buena opción.

El muestreo regular sitúa puntos a intervalos fijos, constituyendo lo que se conoce como una *malla de muestreo*.

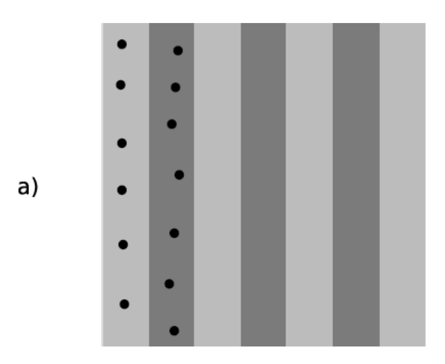
Por último, el muestreo estratificado requiere la presencia de una variable adicional relacionada. Si esta variable se encuentra zonificada, podemos subdividir el muestreo haciendo uso de las distintas zonas. Por ejemplo, si muestreamos una serie de parcelas donde estudiamos la altura media de sus árboles y disponemos de información sobre el tipo de suelo, podemos muestrear por separado para los distintas clases de suelo, ya que este tiene influencia sobre el crecimiento de los pies. Esto aumenta la precisión del muestreo, ya que en lugar de una única población heterogénea, se tienen un numero de subpoblaciones más homogéneas, en virtud de la agrupación según un factor común.

Los diseños pueden combinarse para obtener resultados más satisfactorios. Por ejemplo, puede hacerse un diseño estratificado y en cada uno de los estratos (cada zona definida por la variable adicional) plantear un muestreo regular con un tamaño particular de malla o incluso con un número de puntos a muestrear distinto en cada estrato.

Aplicando conjuntamente todo lo anterior debe tratar de diseñarse un muestreo que cumpla con lo siguiente:

- Explicar la variabilidad de la variable lo mejor posible. Así, en el caso de plantear un muestreo que complemente a uno previo, ya se dispone de información acerca de la variable muestreada. Esta puede emplearse para distribuir adecuadamente los nuevos puntos de muestreo, aplicando que, por ejemplo, se necesitan más muestras en zonas de alta variabilidad.
  El empleo de los denominados *muestreos pilotos* es una herramienta para conocer en primera aproximación las propiedades de la variable a estudiar y diseñar el muestreo acorde con estas.
- Ser representativa. Esta representatividad debe ser tanto en el espacio de atributos como en el espacio geográfico, según puede verse en la figura 14.10

La figura muestra también algunas de las principales debilidades que pueden encontrarse en los distintos métodos de diseño de muestreo. Si la variable exhibe un comportamiento perió-



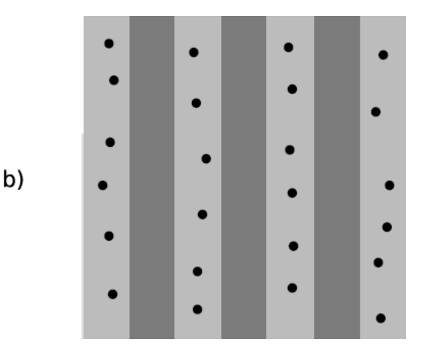


Figura 14.10: El muestreo a) es representativo en el espacio de atributos pero no en el geográfico. El b), sin embargo, es representativo en el espacio geográfico pero no en el de atributos.

dico, un muestreo regular puede enmascarar la existencia de dicha periodicidad y no ser representativo en el espacio de atributos. Ese es el caso del ejemplo a). En general, si existe algún tipo de orden oculto en la variable a analizar, debe tenerse cuidado a la hora de emplear muestreos sistemáticos, para evitar estos fenómenos. En el caso del ejemplo a), aunque no es un muestreo sistemático, la estructura de los puntos muestrales es bastante regular, lo que provoca que no se obtenga la representatividad en el espacio de atributos, dada la estructura periódica de la variable en el eje x.

En el ejemplo b) el muestreo no es representativo del espacio geográfico ya que el diseño deja grandes áreas sin ser muestreadas. Esto puede suceder al emplear muestreos aleatorios, ya que estos, por su propia aleatoriedad, pueden generar estructuras que dejen amplias zonas sin ser muestreadas o con una intensidad de muestreo insuficiente.

14.2.7. Elección del método adecuado

Junto a los métodos de interpolación que hemos visto, que son los más comunes y los implementados habitualmente, existen otros muchos que aparecen en determinados SIG tales como los de *vecino natural* [142], interpolación *picnofiláctica* [143], u otros. Además de esto, cada uno de dichos métodos presenta a su vez diversas variantes, con lo cual el conjunto global de metodologías es realmente extenso. A partir de un juego de datos distribuidos irregularmente, la creación de una malla ráster regular es, pues, una tarea compleja que requiere para empezar la elección de un método concreto. Este proceso de elección no es en absoluto sencillo.

No existe un método universalmente establecido como más adecuado en todas situaciones, y la elección se ha de fundamentar en diversos factores. Al mismo tiempo, un método puede ofrecer resultados muy distintos en función de los parámetros de ajuste, con lo que no solo se ha de elegir el método adecuado, sino también la forma de usarlo. Entre los factores a tener en cuenta para llevar esto a cabo, merecen mencionarse los siguientes:

- **Las características de la variable a interpolar**. En función del significado de la variable, las características de un método pueden ser adecuadas o no. Si, por ejemplo, interpolamos valores de precipitación máxima anual, no es adecuado utilizar aquellos métodos que suavicen excesivamente la superficie resultante, ya que se estarían perdiendo los valores extremos

que, por la naturaleza del valor interpolado, son de gran interés.

- **Las características de la superficie a interpolar**. Si conocemos a priori algunas características adicionales de la superficie resultante, algunos métodos permiten la incorporación de estas características. Por ejemplo, variaciones bruscas en puntos de discontinuidad tales como acantilados en el caso de interpolar elevaciones son aplicables mediante la imposición de barreras con métodos como el de distancia inversa, pero no con otros como el kriging.
- **La calidad de los datos de partida**. Cuando los datos de partida son de gran precisión, los métodos exactos pueden tener más interés, de cara a preservar la información original. Si, por el contrario, sabemos que los datos de partida contienen mucho ruido, aquellos métodos que suavizan el resultado tales como el kriging son preferibles, de cara a atenuar el efecto de dicho ruido.
- **El rendimiento de los algoritmos**. Algunos algoritmos como los basados en distancia son rápidos y requieren un tiempo de proceso aceptable incluso en conjuntos de datos de gran tamaño. Otros, como el kriging, son mucho más complejos y el tiempo de proceso es elevado. A esto hay que sumar la configuración propia del método, con lo que crear una capa ráster con algunos métodos puede llevar mucho más tiempo que con otros y requerir un esfuerzo mayor.
- **El conocimiento de los métodos**. Por obvio que parezca, debe conocerse bien el significado del método para poder aplicarlo. Un método de gran complejidad como el kriging exige una sólida base de conceptos geoestadísticos para su aplicación. Más aún, el elevado número de ajustes que requiere y la alta sensibilidad del método a la variación de estos valores refuerza lo anterior.

  La figura 14.11 muestra tres capas interpoladas mediante kriging a partir de un conjunto de puntos. El empleo de distintos ajustes y variogramas (lógicamente, no todos correctos) da lugar a capas muy diferentes. Si no se emplea correctamente, un método de interpolación puede producir resultados carentes de sentido, que aunque dan lugar a una capa con datos en todas sus celdas, dichos datos pueden no ser adecuados.

  Pese a lo anterior, el kriging es utilizado habitualmente sin considerar estos aspectos, obteniéndose resultados aparentemente correctos pero con frecuencia erróneos. La experiencia del usuario es un elemento clave en la aplicación de métodos de interpolación, y debe ponerse gran énfasis en este aspecto.

  No debe olvidarse tampoco que algunos métodos asumen que se dan ciertas condiciones de los datos de partida, y esto puede no ser cierto, o bien requerirse algún tipo de transformación para que así suceda. Es necesario siempre comprobar que se dan estos supuestos.
- **El uso de la capa resultante**. No es lo mismo utilizar un MDE para crear una vista 3D con una fotografía aérea, que emplearlo para crear una ortofoto. Los requerimientos de calidad en el primer caso son menores, tan solo de tipo visual, y cualquiera de los métodos puede sernos válido. Aplicar una metodología compleja y laboriosa quizás no sea la mejor opción en este caso, y sí lo sea el empleo de una más sencilla.

14.2.8. Elección de las características de la capa resultante

Los métodos que hemos visto en este capítulo no imponen restricciones sobre la distribución o el número puntos de entrada (si bien el kriging, por ejemplo, exige un cierto número de puntos para un ajuste fiable del variograma teórico), ni tampoco sobre las características de la capa rás-

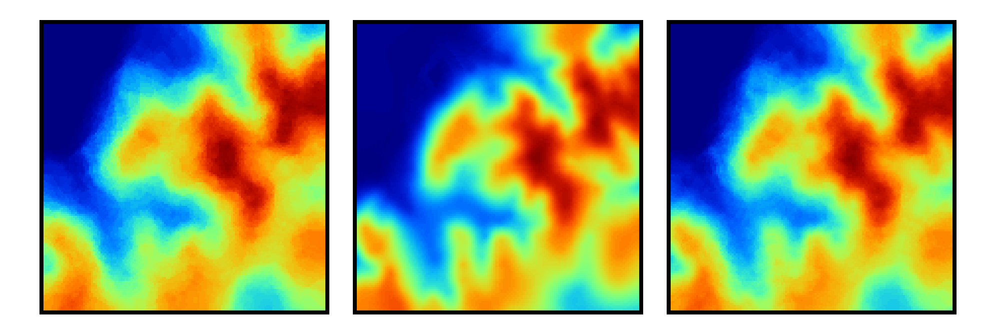


Figura 14.11: Distintos resutlados obtenidos por kriging a partir de un mismo juego de puntos, utilizando diferentes ajustes.

ter resultante. A pesar de ello, resulta claro que existe una relación directa entre ambas, y que la capacidad de un conjunto de valores puntuales para generar una capa ráster es limitada.

En la práctica, a partir de cualquier capa de puntos podemos obtener cualquier capa ráster, ya que pueden siempre calcularse los valores en las celdas. Pero si aplicamos algunos conceptos cartográficos e ideas básicas de teoría de la información, no es difícil darse cuenta de que esto puede ser incorrecto. Existen unas características idóneas para la capa ráster interpolada, fuera de las cuales no es adecuado interpolar.

Vimos en 5.3.1 que la resolución horizontal o tamaño de celda era uno de los parámetros básicos que definían las características de una capa ráster. Existiendo relación entre los puntos de origen y la capa interpolada a partir de ellos, debe existir por igual una relación entre la distribución espacial de los valores puntuales y dicho tamaño de celda, pues ambos elementos definen la precisión con que se recoge la variable estudiada.

Cuando el número de puntos no es suficiente para ello, crear una capa con una resolución alta (tamaño de celda pequeño) equivale a generar un resultado cartográficamente incorrecto. Por el contrario, interpolar con un tamaño de celda demasiado grande supone estar desperdiciando parte de la información de partida, lo cual, en el caso de trabajar sobre un muestreo realizado específicamente para ello, implica un malgasto de medios materiales y humanos en la recogida de datos. La elección del tamaño de celda adecuado es, por tanto, fundamental en el proceso de interpolación.

Los conceptos que vamos a ver a este respecto pueden emplearse tanto para estimar el tamaño de celda de la capa a interpolar, como para ayudar en el diseño del muestreo previo a la creación de esta. Así, si conocemos de antemano (por ejemplo, por el uso que le vamos a dar) las características de la capa que necesitamos generar, podemos diseñar de forma más correcta el muestreo haciendo uso de estas ideas.

[144] argumenta que el tamaño de celda adecuado de una capa ráster es función de la escala, la capacidad de proceso, la precisión posicional, la densidad de muestreo, la correlación espacial existente y la complejidad de la superficie a interpolar. Según estos conceptos, define tamaños mínimos y máximos de celda para cada circunstancia y cada factor considerado de los anteriores. Asimismo, recomienda una solución de compromiso entre los citados valores extremos.

Para la densidad de muestreo, se considera que, por ejemplo para la toma de datos de sue-

lo, por cada centímetro cuadrado del mapa debe existir al menos una muestra [145]. Aplicando los conceptos que vimos en 3.4, se llega a una resolución óptima

$$\Delta s = 0{,}0791\sqrt{\frac{A}{N}} \tag{14.2.11}$$

siendo A el área de la zona a interpolar y N el número de puntos disponibles.

Si, en lugar de lo anterior, aplicamos criterios basados en la disposición geométrica de los puntos, se llega a un valor

$$\Delta s = 0{,}5\sqrt{\frac{A}{N}} \tag{14.2.12}$$

La gran diferencia existente entre este tamaño recomendado y el anterior se debe al hecho de que este último es válido únicamente para patrones de puntos completamente regulares. En el caso de patrones aleatorios o agregados se deben aplicar consideraciones distintas. Por ejemplo, para el caso de un patrón agregado, la distancia media entre puntos es aproximadamente la mitad de la existente en un patrón regular con el mismo número de puntos. Por ello, la anterior fórmula quedaría como

$$\Delta s = 0{,}25\sqrt{\frac{A}{N}} \tag{14.2.13}$$

Aunque resulta imposible dar una cifra exacta para la resolución óptima considerando todos los factores implicados, valores recomendados como los anteriores son de gran utilidad para poder al menos tener una idea del intervalo en el que el tamaño de celda a escoger debe encontrarse. Sin constituir «recetas» infalibles a la hora de elegir un tamaño de celda, constituyen herramientas útiles que deben considerarse antes de interpolar un conjunto de puntos.

14.2.9. Comprobación del ajuste. Validación

Algunos métodos como el kriging dan una estimación del error además de la capa de valores interpolados. En los métodos aproximados tales como el ajuste de funciones polinómicas, puede calcularse el residuo cuadrático según

$$\varepsilon = (\widehat{z} - z)^2 \tag{14.2.14}$$

Es decir, como el cuadrado de la diferencia entre el valor real y el estimado.

En los métodos exactos, por propia definición, este valor es cero, lo cual no quiere decir, lógicamente, que la capa interpolada sea «perfecta» (lo es, pero solo en los puntos de partida), ni que no pueda llevarse a cabo una validación.

Una forma habitual de proceder es, si se dispone de un juego grande de datos (con muchos puntos), «reservar» algunos de ellos para una validación posterior. Se interpola así una capa a partir de la mayor parte de ese conjunto, y se comprueba la diferencia entre los valores estimados en los puntos restantes y los valores medidos en dichos puntos. Los valores estimados en esos puntos no han sido influenciados por los valores reales, ya que no estos se han empleado en la interpolación.

Si no disponemos de un conjunto de datos de validación y el juego de datos disponible es reducido y no podemos separar una parte de él para validar el resultado, podemos llevar a cabo un proceso de validación cruzada. En este proceso, se calculan por interpolación puntos en cada una de las coordenadas de los puntos de muestreo, empleando en cada caso todos los puntos restantes.

De esta forma, obtenemos un conjunto de pares de valores con los valores reales y estimados en cada punto, que podemos mostrar en una gráfica como la de la figura 14.12. De igual mo-

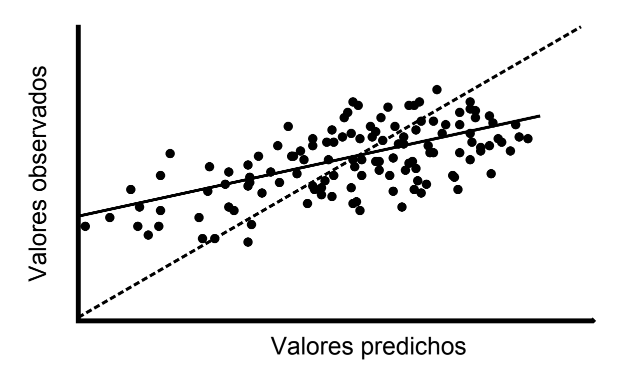


Figura 14.12: Validación cruzada previa de datos puntuales para interpolación.

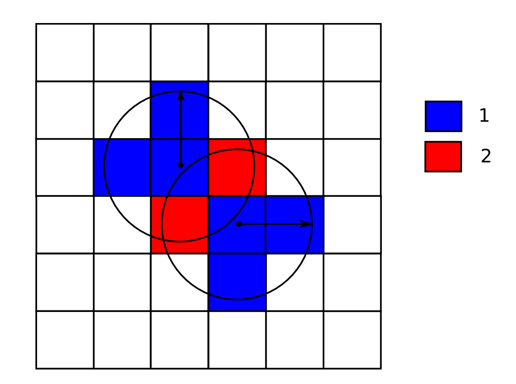


Figura 14.13: Esquema del cálculo de densidades utilizando un área de influencia fija.

do, podemos efectuar un análisis de regresión y obtener la expresión de la recta de ajuste, el coeficiente de correlación de esta o el residuo medio cuadrático. El valor representativo en este caso, no obstante, es la propia expresión de la función. En caso de un ajuste perfecto, la recta debe ser de la forma $y = x$, por lo que resulta interesante representar esta recta ideal para poder comparar.

14.3. Densidad

Además de los métodos de interpolación, hay otro tipo de operaciones que pueden realizarse sobre capas de puntos, y que dan como resultado capas ráster.

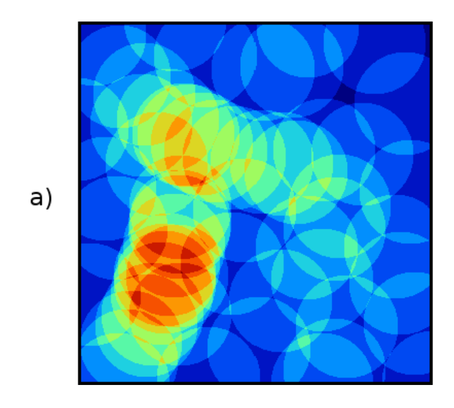
Supongamos el siguiente caso. Se tiene una capa de puntos como la de la figura 14.14 que indica los registros de observación de una determinada especie. Si queremos crear un mapa de densidad de individuos, nos encontramos en un caso similar al que implica la interpolación: la carencia de datos a lo largo de todo el espacio. En este caso, al igual que entonces, podemos aplicar formulaciones específicas para calcular dicha densidad.

Como se vio en el apartado 13.4, la densidad representa la intensidad (propiedad de primer orden) del patrón de puntos, que es una realización concreta del proceso de puntos que lo ha generado.

Por simplicidad, ignoraremos por el momento los atributos asociados a cada punto. Es decir, consideraremos únicamente el dato espacial, que nos dice que en las coordenadas del punto ha sido observada una especie. Asociando a cada punto un área de influencia (por ejemplo, el radio de movimiento conocido de la especie), podemos calcular la capa de densidad de esta sin más que contar en cada celda el número de puntos observados cuyo área de influencia incluye a dicha celda, y dividiendo después por la superficie del área de influencia.

La figura 14.13 muestra un ejemplo sencillo con la áreas de influencia de dos únicos puntos. Donde ambas áreas intersecan, la densidad es, lógicamente, mayor. Las celdas que no están en el área de influencia de ningún punto tienen un valor de densidad nulo.

La elección del área de influencia afecta directamente al resultado obtenido, tal como puede verse en la figura 14.14, que muestra dos capas



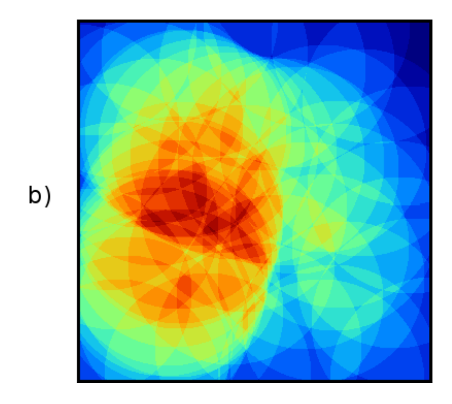


Figura 14.14: Capas de densidad a partir de una capa de puntos. El tamaño del radio de influencia en (b) es el doble que en (a).

de densidad calculadas para distintos radios de influencia.

Se asume por defecto que el área de influencia es circular (mismo alcance de dicha influencia en todas direcciones), y que es la misma para todos los puntos.

En realidad, puede entenderse la creación de una capa de densidades como un histograma bidimensional, siendo por tanto una función de probabilidad.

La pertenencia o no al área de influencia de cada punto es una función discontinua. Para crear capas de densidad con transiciones suaves, pueden utilizarse funciones distintas, de modo que la influencia varíe con la distancia de forma continua.

En estadística, una *función núcleo* o *núcleo (kernel)* es una función de densidad bivariante y simétrica, empleada en la estimación de funciones de densidad de probabilidad de variables aleatorias. El caso anterior en el que todos puntos de la ventana de influencia tienen el mismo peso es un caso particular de núcleo, pero existen muchos otros que dan lugar a estimaciones no discontinuas. La teoría acerca de este tipo de funciones y su uso no se desarrollará aquí, pudiendo consultarse, por ejemplo, en [146].

Un núcleo de uso frecuente es el mostrado en la figura 14.15, denominado *gaussiano*, cuya aplicación resulta en capas suavizadas y continuas (Figura 14.17), y cuya expresión es

$$k(h) = \left(1 - \frac{h^2}{r^2}\right)^2 \qquad \forall h \leq r \qquad (14.3.1)$$

donde h es la distancia al punto y r el radio máximo de influencia.

Como puede observarse comparando las figuras 14.14 y 14.17, la diferencia entre el uso de uno u otro núcleo es notable, pero la utilización de radios distintos, independientemente del método empleado, da lugar a resultados muy diferenciados. El radio de influencia máxima es, pues, el parámetro más importante en la estimación de densidades, y debe seleccionarse de acuerdo con la distribución de los puntos muestrales.

Además de emplear las coordenadas de cada punto para la estimación de densidad, pueden ponderarse estos con los atributos de cada uno de ellos. Así, partiendo de una capa de puntos que represente núcleos poblacionales, para cada uno de los cuales se disponga de su número de habitantes, puede obtenerse una capa de densidad de población. O, en el caso de los puntos que

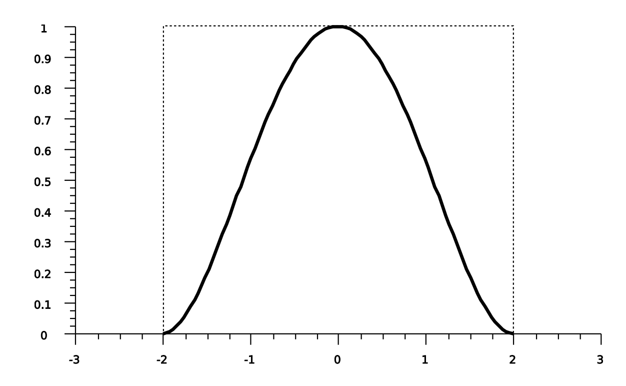


Figura 14.15: Comparación entre una función núcleo constante (en trazo punteado) y un núcleo gaussiano (en trazo continuo) para un radio de influencia $r = 2$.

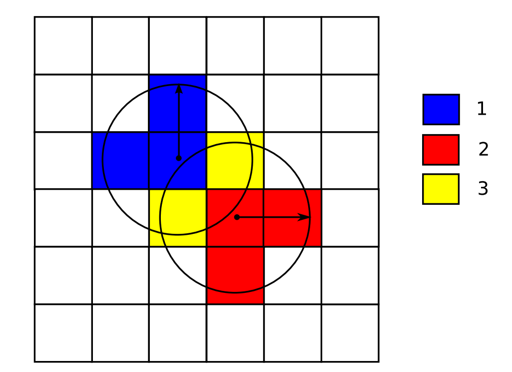
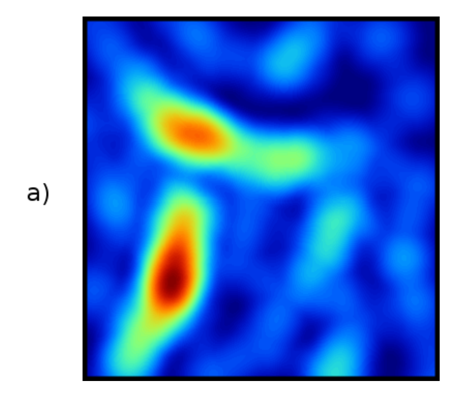


Figura 14.16: Esquema del cálculo de densidades utilizando un área de influencia fija y ponderación según valores. El punto inferior derecho tiene un peso $p = 2$, mientras que el superior izquierdo un peso $p = 1$.



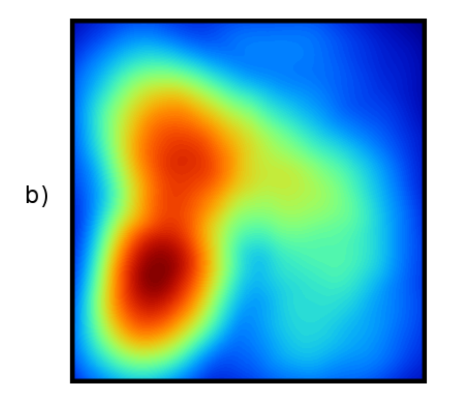


Figura 14.17: Capas de densidad a partir de una capa de puntos mediante nucleo gaussiano. El tamaño del radio de influencia en (b) es el doble que en (a).

representan observaciones de individuos de una especie, recoger en estos puntos no solo el hecho de que se ha producido un avistamiento, sino cuántos de ellos.

La figura 14.16 muestra un ejemplo del cálculo de densidades si se emplea ponderación de los puntos.

En el caso de las observaciones de una especie, la capa de densidad nos da igualmente una medida de la probabilidad de encontrar esa especie en cada celda. El cálculo de estas probabilidades es la base para el desarrollo de modelos predictivos más complejos. Estos modelos incluyen la utilización de variables de apoyo, así como muestreos no solo de presencia, sino también de ausencia.

14.4. Resumen

La información vectorial, en particular la disponible en forma de puntos, puede convertirse en capas ráster a través del uso de métodos de interpolación. Estos permiten calcular valores en puntos no muestreados, a partir de valores re-

cogidos en una serie de localizaciones. De este modo, se puede asignar un valor a cada celda de una capa ráster, y crear una representación continua de la variable recogida en los puntos de muestreo.

Existen muchos métodos de interpolación distintos, entre los que cabe destacar los siguientes por su implementación habitual en los SIG:

- Vecino más cercano
- Ponderación por distancia
- Ajuste de polinomios
- Curvas adaptativas
- Kriging

La elección del método a emplear debe realizarse en función del tipo de datos a interpolar, las características de estos, y los requerimientos y usos previstos de la capa resultante, entre otros factores. Asimismo, es importante elegir una resolución de celda adecuada y, en caso que los puntos de muestreo no vengan dados *a priori*, diseñar un muestreo óptimo.

Empleando métodos de validación y validación cruzada, puede comprobarse la bondad de ajuste de la capa interpolada y la validez de los datos de partida y el modelo empleado.

Junto con los métodos de interpolación, el calculo de densidades permite igualmente la creación de capas ráster a partir de datos puntuales.

Capítulo 15

Álgebra de mapas

15.1. Introducción

Se entiende por *álgebra de mapas*[1] el conjunto de técnicas y procedimientos que, operando sobre una o varias capas en formato ráster, nos permite obtener información derivada, generalmente en forma de nuevas capas de datos. Aunque nada impide que este proceso se lleve a cabo sobre capas vectoriales, se entiende que el álgebra de mapas hace referencia al análisis desarrollado sobre capas ráster, pues estas, por su estructura regular y sus características inherentes, son mucho más adecuadas para plantear los algoritmos y formulaciones correspondientes. Los procedimientos que se aplican sobre información geográfica vectorial se clasifican generalmente dentro de otros bloques de conocimiento, como es por ejemplo el caso de las operaciones geométricas (incluidas dentro de la geometría computacional), cuyos procesos se tratarán en el capítulo 19.

Conviene aclarar que, en la práctica y el uso diario de los SIG, el término *álgebra de mapas* suele utilizarse de forma errónea. Ello es debido a que la gran mayoría de las aplicaciones SIG implementan algún tipo de funcionalidad para la combinación de un número dado de capas ráster, de forma que pueden relacionarse mediante expresiones matemáticas para la obtención de una nueva capa, y esta funcionalidad se designa normalmente como *álgebra de mapas* o en ocasiones *calculadora de mapas*.

La gran utilidad de esta funcionalidad y su uso habitual hacen que se asocie casi exclusivamente con ella el concepto de álgebra de mapas, olvidando —o desconociendo—, que también las restantes funcionalidades de análisis emplean el álgebra de mapas como base fundamental. Es por ello más correcto utilizar en tal caso esa segunda denominación, *calculadora de mapas*, para dichas herramientas, ya que, si bien el uso de estas operaciones entre capas es una aplicación de los conceptos propios del álgebra de mapas, no es la única, y no debe pensarse que todo él es expresable de ese modo.

Como tal, el álgebra de mapas lo forman un conjunto de variables (los mapas), expresiones y funciones, los cuales, a través de una sintaxis adecuada, permiten la obtención de nuevos resultados geográficos. Las funciones que se implementan en las calculadoras de mapas son, como veremos, un subconjunto de las posibles, lo que bien podríamos denominar una *aritmética de mapas*. El concepto de álgebra de mapas, más extenso que el anterior, constituye sin embargo no una herramienta puntual, sino un completo marco de trabajo para el manejo de capas ráster

[1] Aunque esta denominación es la original y más extendida, puede dar lugar a malentendidos. No debe confundirse aquí la idea de *mapa* y pensar que hablamos de un documento cartográfico en su sentido clásico, sino que hablamos de capas, y en particular de capas ráster. Sería más correcto hablar de un *álgebra de capas ráster*, pero se mantiene por costumbre el uso de *álgebra de mapas* para definir al conjunto de técnicas que veremos a lo largo de este capítulo.

y, muy especialmente, su análisis encaminado a la obtención de nuevos resultados.

Si analizamos las practicas geográficas hasta nuestros días, vemos que el álgebra de mapas como proceso de análisis no es algo nuevo. La idea de utilizar mapas existentes para generar otros nuevos o simplemente extraer de ellos resultados cuantitativos es una práctica común desde el mismo momento en que aparece la cartografía moderna. Sin embargo, es con la aparición de los Sistemas de Información Geográfica y la posibilidad de procesar los datos geográficos en un entorno informatizado cuando se dota de formalismo a estos planteamientos y se define con rigor el conjunto de herramientas de análisis.

La definición actual del álgebra de mapas la debemos a Dana Tomlin [1], quien estableció la división principal de funciones y dio cuerpo a la disciplina, sentando así las bases para lo que es hoy en día la manera habitual de proceder en el análisis de capas geográficas ráster. Posteriormente, otros autores han desarrollado generalizaciones y extensiones de las ideas de Tomlin, pero son estas las que, desde el punto de vista práctico, cubren la mayor parte de casos posibles y resultan de utilidad directa para el análisis habitual.

Más allá de lo que se desarrolla en este capítulo, las ideas que veremos a continuación se van a repetir de forma continua en la mayoría de capítulos de esta parte, pues constituyen la base formal y conceptual de los análisis que siguen en dichos capítulos.

Para comenzar el estudio del álgebra de mapas, veamos unos casos prácticos que nos servirán para introducir con posterioridad los elementos básicos. En primer lugar, consideremos la siguiente expresión, que define la bien conocida Ecuación Universal de Pérdidas de Suelo (USLE) [147]

$$A = R \cdot K \cdot LS \cdot C \cdot P \qquad (15.1.1)$$

donde A representa las pérdidas totales en toneladas por hectárea y año, y los factores R, K, LS, C y P representan la influencia de los diversos factores (agresividad del clima, tipo de suelo, topografía, uso de suelo, y prácticas de conservación) sobre dichas pérdidas. La USLE representa, por tanto, un modelo sencillo que combina cinco variables, todas ellas susceptibles de ser recogidas en las correspondientes capas ráster. Si extendemos el cálculo puntual de la variable A a todos los puntos de la zona estudiada, obtendremos una nueva capa de dicha variable, evaluando la anterior expresión para cada una de las celdas de esas capas ráster (Figura 15.1).

La discretización del espacio que implica la representación en formato ráster de un parámetro sobre una superficie dada, hace que ese análisis de «todos los puntos» se materialice en un análisis de todas las celdas que contiene la nueva capa a obtener. Esto nos hace ver la idoneidad de este formato para trasladar las ideas matemáticas de modelos sencillos como la USLE a un álgebra de mapas donde estos modelos no se apliquen en un emplazamiento concreto, sino en toda una región de interés a tratar.

En este caso, basta con aplicar la expresión mostrada en la figura 15.1 celda a celda, y obtener tantos valores resultantes como celdas haya, que conformarán una nueva capa con la distribución espacial de las pérdidas de suelo correspondientes.

Consideremos ahora que, junto con el mapa de pérdidas de suelo obtenido según lo anterior, disponemos de una división en subcuencas de la zona de estudio. Puede resultar interesante asociar a cada una de las unidades hidrológicas un valor relacionado con los valores de pérdidas de suelo que se dan en ella. Por ejemplo, la media

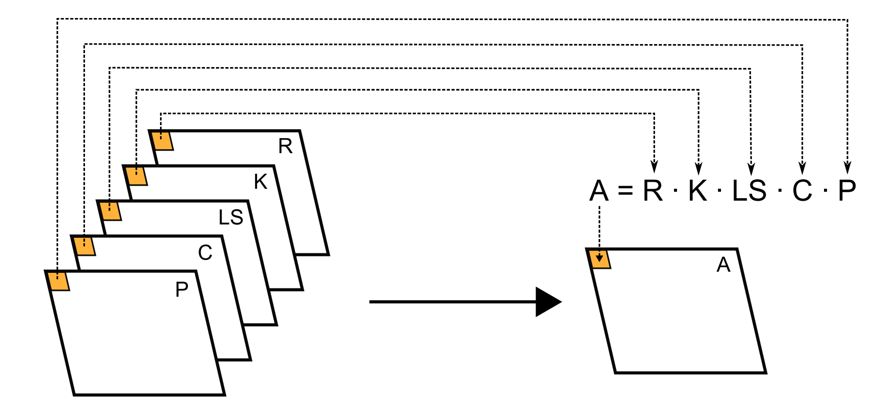


Figura 15.1: Puede extenderse una formula algebraica tal como la de la Ecuación Universal de Pérdidas de Suelo (USLE) a un conjunto de capas. Basta operar celda a celda con las capas que contienen las variables implicadas, para obtener una nueva capa resultante con el conjunto de valores calculados.

de los valores de pérdidas de suelo de todas las celdas de la subcuenca. Partiendo de estos dos mapas, podemos obtener un tercero que nos indique las pérdidas medias en cada cuenca, de interés sin duda para localizar las unidades que puedan presentar mayores problemas de erosión (Figura 15.2).

La base conceptual en este caso vemos que es distinta al primer supuesto, ya que no extendemos ningún modelo puntual aplicándolo en todas las celdas con distintos valores de entrada para cada una de ellas, sino que estas van a contener un valor medio asociado al total de celdas incluidas en su misma subcuenca. Esta forma de proceder ya no se corresponde con el caso anterior y no puede ser llevada a cabo con las funcionalidades de *calculadora de mapas* que citábamos como habituales en los SIG, ya que no se da un análisis entre capas celda a celda, sino un análisis dentro de la propia capa con otras celdas relacionadas a través de esa pertenencia a la misma unidad hidrológica. No obstante, como veremos, este tipo de funciones también forman parte del álgebra de mapas.

Por último, supongamos que es de interés estudiar la distribución de valores de esas pérdidas de suelo. Un histograma de frecuencias resultaría de gran utilidad. En este caso, el análisis trabaja una vez más sobre la capa de pérdidas de suelo, pero no genera nuevos resultados geográficos. Aun así, la generación de nuevas capas no es una condición básica del álgebra de mapas, ya que sus procesos pueden tener como resultado elementos muy diversos.

Todos estos ejemplos son parte de los resultados que pueden obtenerse utilizando los distintos componentes del álgebra de mapas, que a continuación detallaremos.

15.2. Tipos de funciones en el álgebra de mapas

Las funciones son el elemento principal del álgebra de mapas. Cuatro son los tipos principales de funciones que podemos definir, agrupadas según la forma en que toman la información necesaria para su cálculo de entre la contenida en todas las celdas de las capas de origen.

- **Local**. El valor en cada celda de la capa resultante es función únicamente de los valores en esa misma celda en las capas de partida.



Figura 15.2: Valores de pérdidas de suelo por subcuencas.

- **Focal**. El valor en cada celda de la capa resultante es función del valor en dicha celda y en las situadas en un entorno definido alrededor de la misma.
- **Zonal o regional**. El valor en cada celda de la capa resultante es función del valor de todas las celdas conectadas a esta que presentan un mismo valor para una de las capas de entrada (pertenecen a la misma clase que esta).
- **Global**. El valor resultante de la función es obtenido a partir de todas las celdas de la capa.

La combinación de distintas funciones y de enfoques variados da lugar a un enorme conjunto de operaciones de análisis basadas en el álgebra de mapas así definido. Este conjunto es el que dota de toda su potencia a los SIG como herramientas de análisis del medio, y permite extraer de los datos geográficos en formato ráster toda la información que realmente contienen.

15.2.1. Funciones locales

Las funciones locales asignan valores a una celda en base a los valores que esa misma celda presenta para cada una de las capas de entrada, operando con estos de una forma u otra. Es decir, el valor resultante para una localización dada es función exclusivamente de lo que se encuentra en dicha localización, no dependiendo en modo alguno de otras localizaciones (otras celdas) (Figura 15.3).

Las funciones locales son las que utilizamos cuando empleamos esa anteriormente citada *aritmética de mapas*, tal y como veíamos por ejemplo en el caso de la USLE. Aunque otro tipo de funciones también operan con varias capas, la combinación de una serie de ellas suele llevarse a cabo con funciones locales, que calculan los valores para cada punto de acuerdo con los valores de dichas capas en ese punto. El ejemplo de la USLE es una función sumamente sencilla, que tan solo multiplica las capas de partida, pero pueden elaborarse funciones más complejas utilizando todos los operadores disponibles, que estudiaremos más adelante.

Una función de tipo local puede ser también aplicada sin necesidad de tener una serie de capas, sino con una única capa de partida. Por ejemplo, un cambio de unidades es una función local, ya que cada uno de los valores expresados en las unidades de destino solo depende del valor expresado en las unidades de origen en cada propia celda. Si una capa conteniendo elevaciones expresadas en metros la multiplicamos por 100, obte-

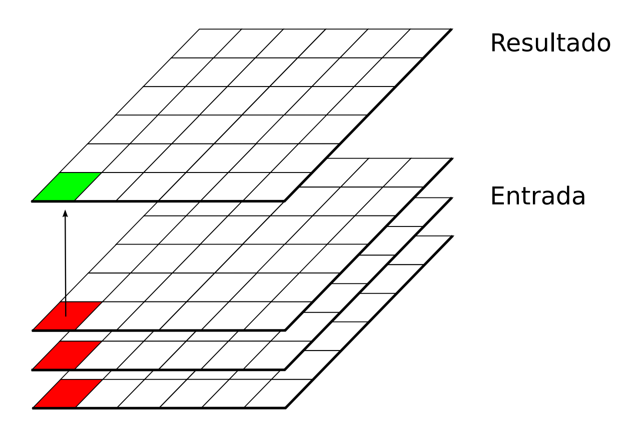


Figura 15.3: Las funciones de análisis local analizan los valores de una celda concreta en una serie de capas para obtener el valor resultante. En rojo, celdas de partida analizadas. En verde, celda que recoge el resultado.

nemos una nueva capa con valores de elevación en centímetros, habiendo aplicado una función local para realizar la conversión.

De igual modo, convertir los valores de una capa de pendientes de radianes a grados requiere multiplicar sus valores por $180/\pi$.

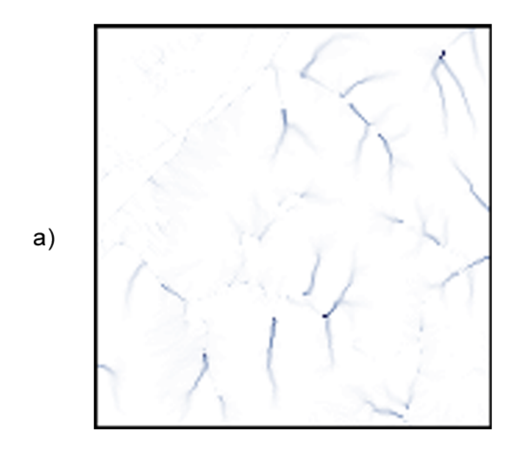
Además de operar con escalares, podemos aplicar también funciones matemáticas. Por ejemplo, las capas que presentan un gran coeficiente de variación (elevada varianza en relación a la media de los valores de la capa) muestran mucha más información si aplicamos una transformación logarítmica. En la figura 15.4 vemos una capa de área acumulada a (este concepto hidrológico se explicará en el apartado 16.6.3) junto a otra que representa el parámetro $\log(a)$, apreciándose claramente la diferencia entre ambas en cuanto a la riqueza de información visual que aportan. Esta transformación logarítmica es una función local dentro del álgebra de mapas.

Una aplicación común de una función local con una sola capa la encontramos en la normalización de valores. En ocasiones, antes de efectuar un proceso que englobe a varias capas (por ejemplo, otra función de análisis local pero multicapa), es necesario homogeneizar estas de modo que todas se hallen en un mismo rango de valores. Este proceso se denomina *normalización*. Es habitual que el rango común sea el intervalo $(0,1)$, para lo cual se aplica a cada capa la función local definida por la siguiente expresión:

$$y_{ij} = \frac{x_{ij} - x_{min}}{x_{max} - x_{min}} \tag{15.2.1}$$

donde y_{ij} es el valor normalizado en la celda ij, x el valor de esa celda en la capa inicial y x_{min} y x_{max}, respectivamente, los valores mínimo y máximo de la variable en el conjunto de celdas de la capa.

Si en vez de aplicar una función matemática sobre el valor de cada celda, aplicamos una serie de criterios referidos a dicho valor mediante operadores de comparación, podemos llevar a cabo un proceso de clasificación. De este modo, podemos obtener a partir de una capa continua una capa discreta en la que las celdas serán clasificadas en grupos según el valor de la variable de partida, o bien reconvertir una clasificación ya existente en otra de acuerdo a unas condiciones establecidas. La figura 15.5 muestra una clasificación de los valores de pendiente en clases, práctica habitual en muchas disciplinas a la hora



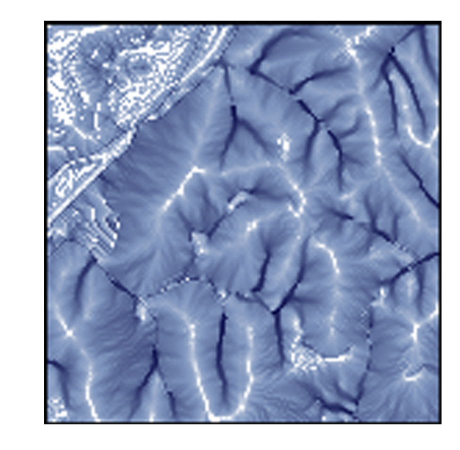


Figura 15.4: La aplicación de una transformación logarítmica modifica la representación visual de una capa. a) Capa original, b) capa tras trasformación logarítmica.

de trabajar con este parámetro. Para ello se ha utilizado el siguiente criterio.

$$s' = \begin{cases} 1 & \text{si } s \leq 5 \\ 2 & \text{si } 5 < s \leq 10 \\ 3 & \text{si } 10 < s \leq 20 \\ 4 & \text{si } s > 20 \end{cases} \tag{15.2.2}$$

donde s' es la clase de pendiente, y s el valor de la pendiente en porcentaje.

También es posible reclasificar capas que ya contienen información categórica, sustituyendo los valores de una clase por un nuevo valor. Puede utilizarse para crear clasificaciones menos detalladas, agrupando clases similares en una única.

Dentro de las funciones locales sobre una única capa, podemos considerar como un caso particular la generación de nuevas capas «desde cero», es decir, sin basarnos en los valores de ninguna capa previa. Por ejemplo, crear una capa de valor constante k o una capa con valores aleatorios dentro de un intervalo definido. En este supuesto, se toma de la capa origen solo su extensión y tamaño de celda, pero los valores son generados sin basarse en los existentes en ella.

Cuando las funciones locales se aplican a varias capas, la forma de combinar estas es muy variable. Junto a las operaciones que ya hemos visto, podemos utilizar algunas otras, y de modos igualmente variados. El conjunto de ellas lo dividimos en los siguientes grupos:

- **Operadores aritméticos**. Para formar expresiones con las distintas capas tales como la ecuación USLE que ya conocemos.
- **Operadores lógicos**. Pueden tomarse los valores de las capas como valores booleanos (1 o 0, verdadero o falso), o aplicar expresiones lógicas de tipo *pertenece al conjunto* u operadores de comparación, entre otros.
- **Parámetros estadísticos**. Por ejemplo, el valor mayor de entre las todas las capas. También pueden recogerse otros como el orden del valor de una capa dentro de la serie ordenada de valores en todas las capas, el código de la capa donde aparece ese valor mayor, o el numero de capas con valores iguales a uno dado. La figura 15.6 muestra algunos ejemplos simples basados en estas ideas.

Comenzando por el supuesto más sencillo de utilizar únicamente dos capas, podemos aplicar

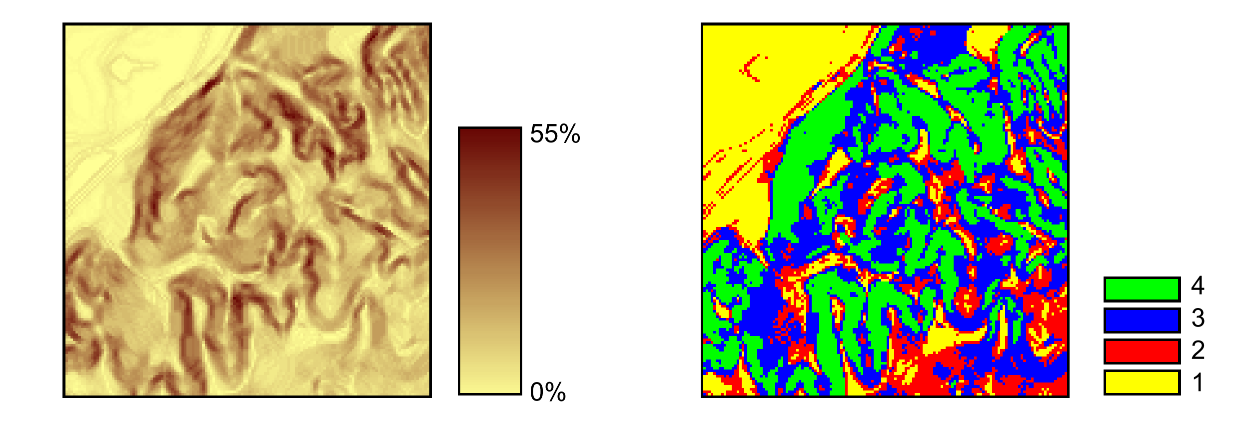


Figura 15.5: Mediante una función local de reclasificación, podemos convertir una capa de valores continuos en una capa de clases con información discreta.

operadores lógicos tales como $<, >, \neq, =, \leq$ o $\geq$. Por ejemplo, con dos capas con información categórica de usos de suelo correspondientes a dos fechas distintas, el operador de desigualdad nos servirá para detectar en la nueva capa resultante aquellas celdas donde el uso de suelo haya cambiado.

Cuando tenemos un conjunto mayor de capas, podemos aplicar los operadores anteriores, e incluso combinar operadores de varios grupos distintos de entre los anteriores. Por ejemplo, la técnica conocida como *Ordered Weighted Average* (OWA) [148], aplica una media ponderada de las capas de la forma

$$y = \sum_{i=1}^{n} z_i k_i \; ; \; k_i \in \mathbb{N} \tag{15.2.3}$$

siendo n el número de capas y z_i el valor i–esimo de los de las distintas capas, ordenados estos en orden ascendente. El valor i_1 sería el más pequeño de todas las capas en la celda problema, y i_n el mayor. Esto hace que el valor z_i que se multiplica por k_i no esté siempre asociado a una capa fija (i no representa a una capa), sino a una posición dentro de la lista ordenada formada por los valores de todas las capas para cada celda.

La aplicación del OWA puede verse como un uso combinado de una función de análisis local de tipo estadístico que se encarga de generar nuevas capas con los valores i-esimos, y la posterior aplicación de una operación aritmética. Esta última ya se aplicaría de la forma habitual, pero sobre las capas provenientes de la primera operación, no sobre las originales.

Aunque si trabajamos con capas de tipo categórico carece de sentido desde un punto de vista conceptual el operar aritméticamente con valores que identifican una clase, las operaciones aritméticas nos pueden servir de igual modo en este caso para obtener nuevas capas. Una función local aritmética nos sirve como herramienta para realizar algunas tareas, entre ellas una habitual como es combinar en una sola capa dos clasificaciones distintas.

Partiendo de una capa de usos de suelo y una de tipos de suelo, podemos obtener una nueva clasificación que combine ambas (Figura 15.7). Un proceso como se realiza, por ejemplo, para el cálculo del Número de Curva [149], una variable

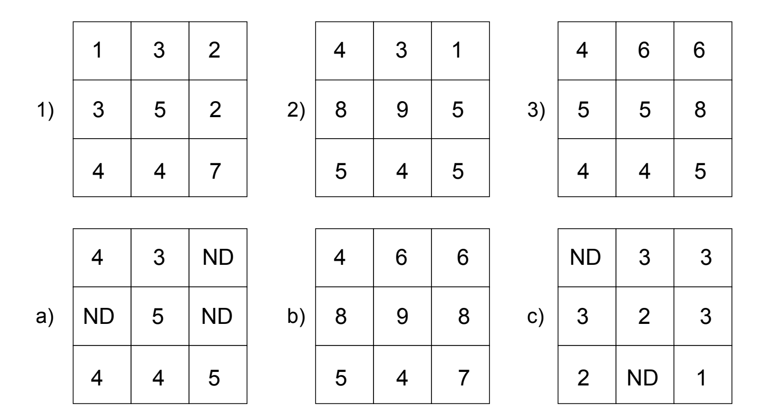


Figura 15.6: Algunos ejemplos sencillos de análisis local estadístico con múltiples capas. En la fila superior, capas de origen y su numeración correspondiente. En la inferior, de izquierda a derecha: valor más frecuente (mayoría), valor máximo, y capa de máximo valor. *ND* indica celdas sin datos en las cuales no puede establecerse un valor resultante por no estar este bien definido.

hidrológica que permite calcular la generación de escorrentía a partir de una precipitación dada. La clase de Número de Curva se asigna en función del uso y el tipo de suelo. El proceso es, en realidad, una intersección de las zonas definidas por cada capa.

Para efectuar esta intersección, debemos en primer lugar reclasificar las capas de inicio de tal modo que un valor en la capa resultante defina unívocamente una única combinación de estas. Después, operaremos con las capas reclasificadas, eligiendo un operador que nos permita mantener esa correspondencia biunívoca entre pares de valores de origen y valor resultante.

Una forma de hacer esto, suponiendo un número de clases m en la primera clase y n en la segunda, es reclasificar la primera de ellas mediante la expresión

$$c'_i = k^i \; ; \; k \in \mathbb{N}, i = 1 \ldots m \qquad (15.2.4)$$

donde c'_i es el nuevo valor a asignar a la clase i–ésima. Es decir, se asignan potencias sucesivas de un valor natural. De igual modo, se asignan los nuevos valores a la segunda capa siguiendo la progresión de potencias, de la forma

$$c'_i = k^{i+m} \; ; \; k \in \mathbb{N}, i = 1 \ldots n \qquad (15.2.5)$$

Con las capas anteriores, basta sumarlas para obtener una nueva en la que el valor de cada celda nos define inequívocamente a partir de qué valores originales se ha calculado.

Para ver un ejemplo que sea manejable, en la figura 15.8 se muestran dos capas con su valores originales, las capas reclasificadas según el esquema anterior, y la capa resultante.

Los SIG más comunes incorporan entre sus elementos funciones que simplifican este proceso y hacen innecesario operar de este modo, por lo que no es corriente tener que aplicar estos razonamientos manualmente. No obstante, resulta de interés el mostrar estas técnicas para estimular y desarrollar la capacidad de razonar espacial y numéricamente en base a los conceptos del álgebra de mapas, conociendo estos con detalle.

Por ultimo, para concluir esta sección es interesante señalar que la gestión de valores que

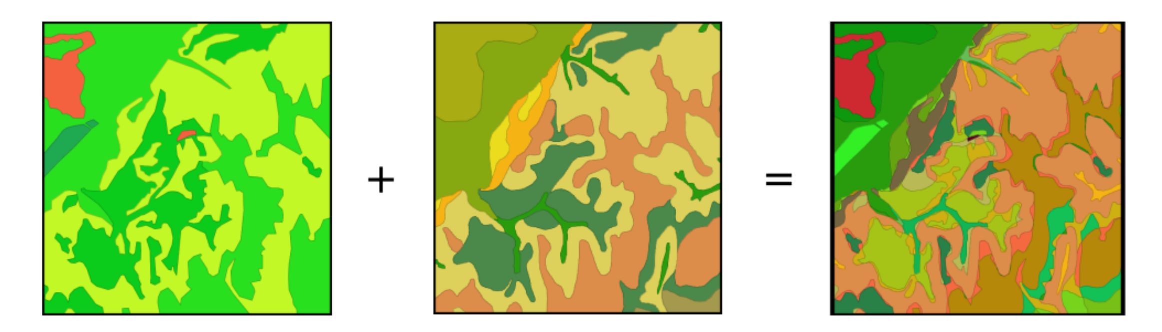


Figura 15.7: Combinación de dos capas categóricas para crear una nueva clasificación que une ambas.

representan la ausencia de datos es un aspecto importante en el empleo de operadores aritméticos en funciones locales. En general, se adopta como práctica habitual el que una operación aritmética entre celdas de varias capas devuelva un valor de sin datos siempre que alguna de las celdas implicadas carezca de datos (es decir, tenga un valor de sin datos). Dicho de otro modo, la presencia de un valor de sin datos en la operación hace que la celda resultante reciba automáticamente también valor de sin datos, particularmente el establecido para la capa resultante.

Esta forma de proceder, además de dar un resultado coherente con los datos de entrada, puede utilizarse como herramienta para, aplicando inteligentemente capas con zonas sin datos, preparar las capas de entrada de cara a su uso en otros análisis. Ese es el caso de la creación de *máscaras*, que nos permiten restringir la información de la capa a una parte concreta de la misma. La figura 15.9 muestra cómo un modelo digital del terreno se recorta para contener información únicamente dentro de una zona definida, en este caso todas las celdas situadas a más de 180 metros de elevación.

Para realizar el recorte, la capa que define la zona de interés contiene valor 1 en las celdas interiores y el valor de sin datos correspondiente en las exteriores. Al multiplicarlo por el modelo digital del terreno, el resultado es la propia elevación en las interiores, y el valor de sin datos en las exteriores, ya que una de las capas no tiene datos suficientes para poder generar otro resultado.

También veremos más adelante que ese uso de mascaras tiene su equivalente vectorial, existiendo una operación de recorte para capas de datos vectoriales.

15.2.2. Funciones focales

Las funciones de análisis focal operan sobre una sola capa de datos, asignando a cada celda un valor que deriva de su valor en la capa de partida, así como de los valores de las situadas en un entorno inmediato de esta (Figura 15.10). La función focal queda así definida por las dimensiones y forma del entorno a considerar, así como por la función a aplicar sobre los valores recogidos en este.

A diferencia de las funciones locales, las focales no se aplican sobre varias capas, ya que la información necesaria se extrae de la vecindad de cada celda, dentro de la propia capa de partida.

Las funciones focales más habituales emplean un entorno cuadrado 3×3 centrado en la celda, que se va desplazando por la capa de tal modo que todas las celdas van siendo designadas como celdas centrales, y se calcula así un nuevo valor para todas ellas. Este entorno de celdas a considerar se denomina frecuentemente *ventana de análisis*.

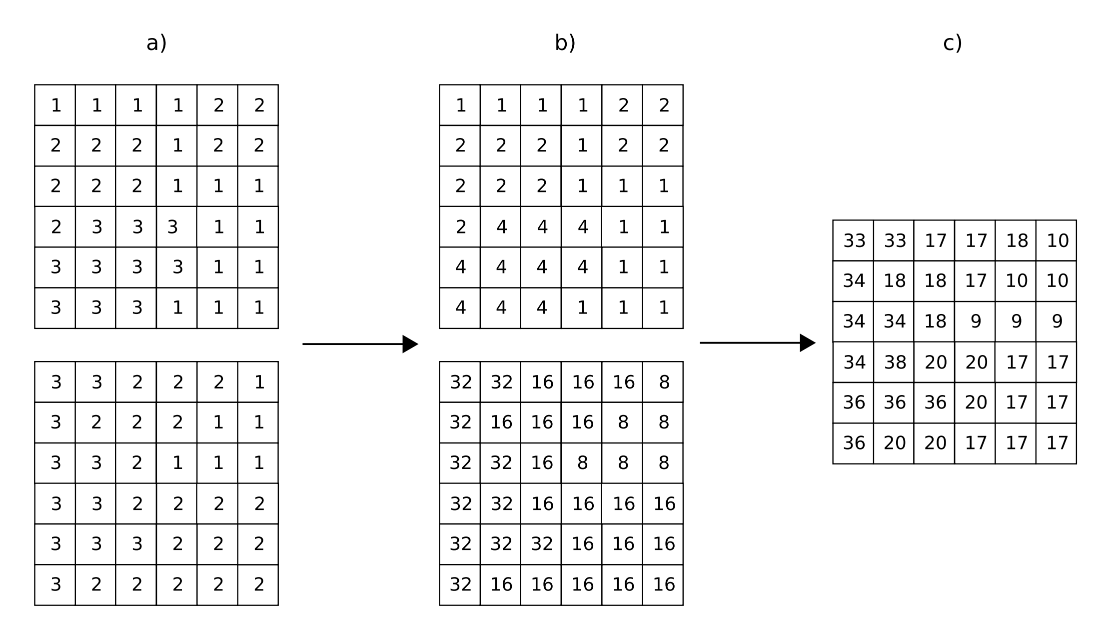


Figura 15.8: Ejemplo de combinación de dos capas categóricas. a) capas originales, b) capas reclasificadas, c) resultado.

Para definir las operaciones sobre esta ventana, es frecuente introducir una notación como la siguiente con el fin de simplificar las expresiones.

z_1	z_2	z_3
z_4	z_5	z_6
z_7	z_8	z_9

siendo z_5 la celda central, la cual recibirá el valor resultante de la operación efectuada. Puesto que los análisis focales basados en esta ventana tipo son habituales, haremos uso de esta notación en diversos puntos dentro de esta parte del libro.

Aunque menos frecuentes, pueden utilizarse ventanas de tamaño mayor, $n \times n$, siendo n un valor impar para que de este modo exista un celda central. De otro modo, la ventana no podría quedar centrada sobre la celda a evaluar, sino desplazada. De igual forma, la ventana no ha de ser necesariamente cuadrada, y otras formas distintas son aplicables. La figura 15.11 muestra algunas de las más comunes, todas ellas también aplicables a distintos tamaños.

Con los valores de las celdas contenidas en la ventana de análisis pueden realizarse operaciones muy diversas, entre las que cabe citar las siguientes:

- **Cálculo de descriptores estadísticos**. Los más habituales son la media, la mediana, los valores extremos o el rango de valores. Para el caso de valores discretos, son comunes parámetros como el número de clases (número de celdas con distinto valor) dentro de la ventana de análisis.
- **Combinaciones lineales**. De acuerdo con expresiones de la siguiente forma:

$$y = \frac{\sum_{i=1}^{n} z_i k_i}{\sum_{i=1}^{n} k_i} \ ; \ k_i \in \mathbb{R} \qquad (15.2.6)$$

Este grupo particular de operaciones se conocen como *convoluciones*, y son la base para una larga serie de procedimientos muy comunes en el



Figura 15.9: Recorte de una capa empleando una mascara con celdas sin datos. La rampa de colores se ha variado en la capa recortada para ajustarse al nuevo rango de valores de esta.

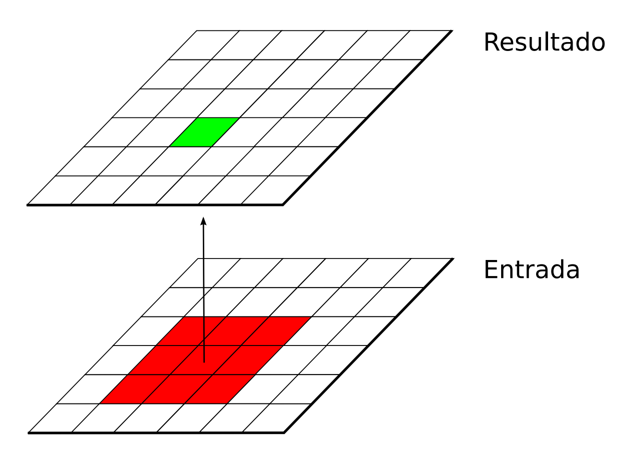


Figura 15.10: Las funciones de análisis focal analizan los valores de una celda y las situadas en un entorno de esta para obtener el valor resultante. En rojo, celdas de partida analizadas. En verde, celda que recoge el resultado.

tratamiento de imágenes digitales. Es habitual expresar el conjunto de valores k_i también en forma de ventana, siendo esta, lógicamente, coincidente en sus dimensiones con la de análisis. Por ejemplo, como en el siguiente caso:

1/9	1/9	1/9
1/9	1/9	1/9
1/9	1/9	1/9

Este conjunto de valores de k_i así expresados se conoce comúnmente como *núcleo* o *kernel* de la convolución. Nótese que el núcleo anterior se corresponde con el cálculo de la media aritmética, pudiendo expresarse este descriptor estadístico como una combinación lineal de los valores de la ventana, a través de un núcleo.

- **Operaciones matemáticas de forma general**. No necesariamente combinaciones lineales, aplican operadores más complejos a los valores de la ventana.
- **Clasificaciones**. En función de la configuración de los valores dentro de la ventana clasifican la celda en una serie de posibles grupos, de acuerdo con unas reglas definidas. El resultado es una capa de información discreta, frente a

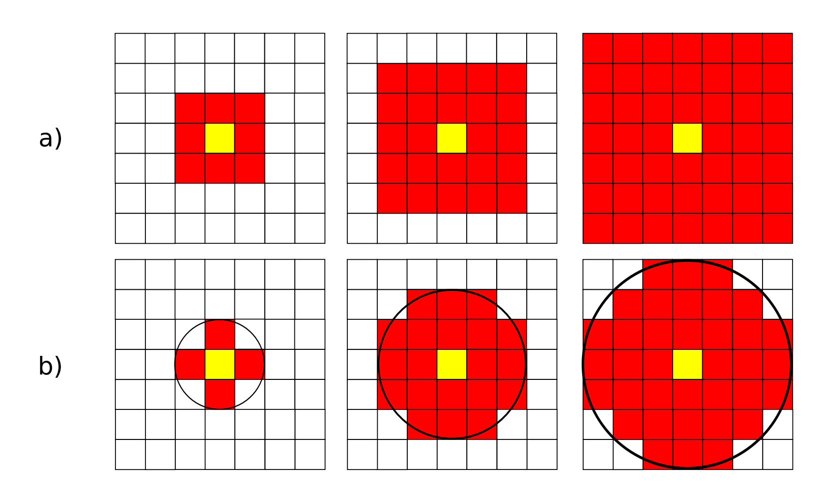


Figura 15.11: Algunos de los tipos de ventana de análisis más comunes en distintos tamaños. a) cuadrada, b) circular

las anteriores que producen capas continuas. Un ejemplo de esto lo encontramos en la clasificación de formas de terreno, la cual veremos en el apartado 16.8, o en la asignación de direcciones de flujo según el modelo D8 (16.6.1).

Algunas de las funciones anteriores se han de definir de forma específica para un tamaño y forma de ventana dado, mientras que otras, como el caso de los descriptores estadísticos, pueden definirse de forma genérica. La diferencia estriba en que en estos la posición del valor dentro de la ventana de análisis no es relevante, mientras que para otras funciones sí lo es.

El resultado de un operador de análisis focal no ha de ser necesariamente un valor que se sitúa en la celda central de la capa resultante una vez ha sido calculado. Por ejemplo, y relacionado con lo comentado en el párrafo anterior, [150] propone un operador que, evaluando los valores dentro de la ventana de análisis, modifique la capa de salida no en la celda central, sino en aquella.que cumpla una condición dada. Por ejemplo, aquella que contenga el valor máximo de entre todas las de la ventana. Lo importante en este caso no es el valor, sino sobre qué celda se sitúa.

Con respecto al tamaño de la ventana de análisis, debe mencionarse que la utilización de uno u otro tiene dos consecuencias directas: por un lado el proceso es más costoso en términos de tiempo a medida que aumentamos la ventana, ya que el número de celdas a analizar es mayor. Por otro, el resultado puede diferir notablemente, y es necesario tener en cuenta el significado del parámetro a calcular para establecer unas dimensiones correctas —dimensiones en unidades reales, no en número de celdas— de la ventana. La figura 15.12 muestra cómo el efecto de un filtro de media, el cual produce un desenfoque de la imagen, se hace más patente a medida que empleamos ventanas de análisis mayores.

En [151] puede encontrarse información adicional sobre la noción de escala de análisis —especialmente para el caso de análisis del terreno— y otros conceptos íntimamente relacionados con la elección de un tamaño de ventana. En el apartado 16.8 veremos un análisis particular en el que la elección del tamaño de ventana es particularmente importante.

Con independencia de dicho tamaño de ventana, siempre vamos a tener algunas celdas para

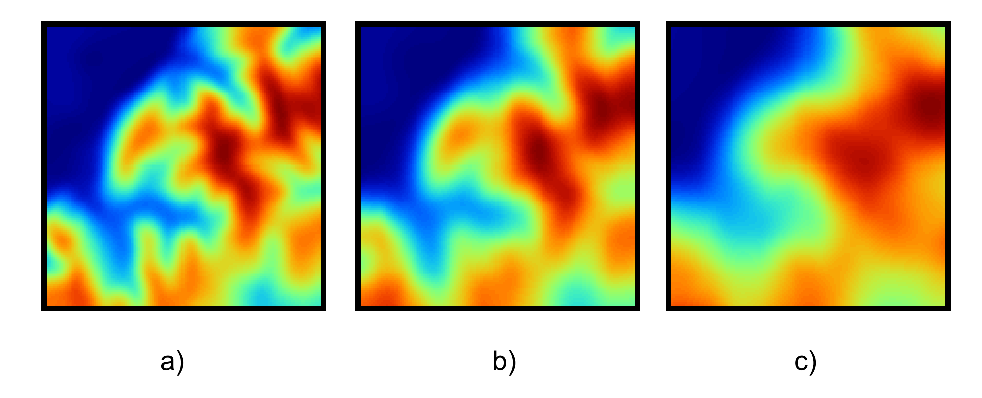


Figura 15.12: Resultados de un filtro de mediana sobre una imagen para distintos tamaños de ventana. a) 5×5, b) 10×10 c) 20×20

las que esta no va a poder definirse en su totalidad. Estas celdas son las situadas en los bordes de la capa, ya que en su caso siempre habrá algunas celdas de la ventana que caigan fuera y para las cuales no tengamos un valor definido (Figura 15.13). En este caso, debe o bien definirse una nueva formulación para estas celdas de borde, o trabajar únicamente con las celdas interiores a la capa, o directamente asignar un valor de sin datos a la capa resultante, indicando que no puede evaluarse el parámetro en ausencia de algún dato. El optar por una u otra alternativa será función, como ya vimos antes, de si el valor resultante depende o no de la posición de los valores de partida.

Para el caso de una media aritmética, si de los nueve valores de la ventana habitual solo tenemos, por ejemplo, seis, podemos operar con ellos y asumir que el resultado será satisfactorio. En el caso de asignar direcciones de flujo, sin embargo, los valores pueden ser erróneos, ya que tal vez el flujo se desplace hacia las celdas fuera de la capa, pero al faltar la información de estas, no será posible hacer tal asignación. Una práctica recomendable en cualquier caso es no limitar la extensión de la capa a la mínima que englobe el área del territorio que queramos estudiar, sino tomar una porción adicional alrededor para que estos efectos de borde no tengan influencia sobre nuestro estudio.

15.2.3. Funciones zonales o regionales

Las funciones de análisis zonal asocian a cada celda valores relativos no a ella misma ni a un entorno fijo alrededor, sino a la clase a la que dicha celda pertenece (Figura 15.14). Se necesita, por tanto, una capa de apoyo que contenga la pertenencia de cada celda a una u otra clase, ya que la utilización de una celda en el análisis no se establece por posición, como en los casos anteriores, sino por valor. Esta capa es de tipo discreto y representa una teselación del territorio en un número definido de clases.

Lo habitual es emplear esta capa de clases en conjunción con otra, ya sea de valores continuos o discretos, y extraer de esta segunda los valores a utilizar para definir el valor representativo de cada clase. Ese es el caso del ejemplo propuesto al principio del capítulo, donde se utiliza el mapa de pérdidas de suelo para asignar los valores correspondientes a cada subcuenca. En este caso, las clases vienen definidas por las subcuencas.

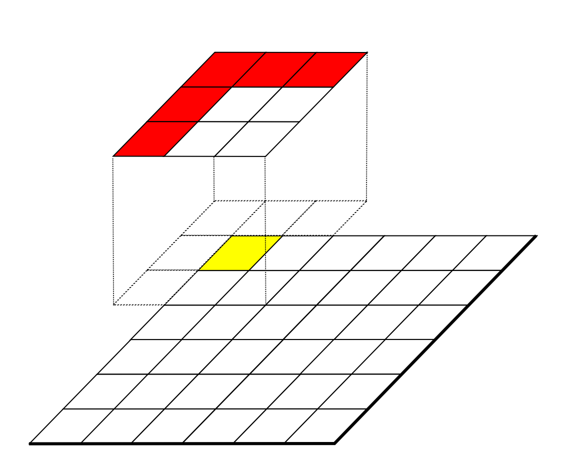


Figura 15.13: En las celdas de borde la ventana de análisis no puede definirse en su totalidad. Las celdas en rojo representan celdas fuera de la capa para las cuales no existe información

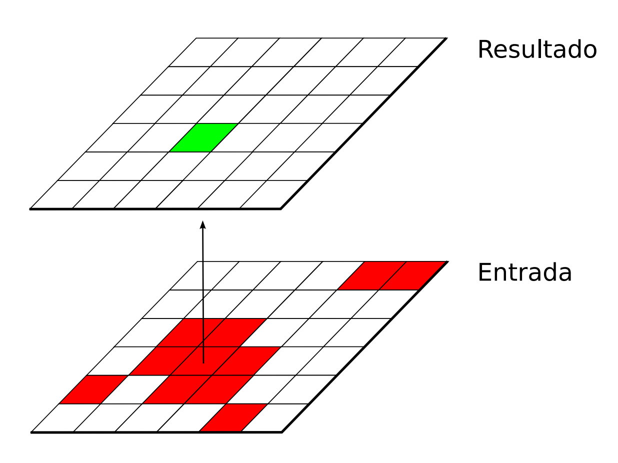


Figura 15.14: Las funciones de análisis zonal analizan los valores de todas las celdas asociadas a una misma clase cada para obtener el valor resultante. En rojo, celdas de partida analizadas. En verde, celda que recoge el resultado.

La definición del conjunto de celdas relacionadas con una dada puede realizarse de dos formas distintas (Figura 15.15):

- Todas las celdas con el mismo valor que la celda problema, conectadas por contigüidad con esta.
- Todas las celdas con el mismo valor que la celda problema presentes en la capa, con independencia de su conexión.

En el caso de las pérdidas por subcuencas, calculábamos con los valores del conjunto de celdas pertenecientes a cada clase su media aritmética, pero pueden aplicarse igualmente diversos descriptores estadísticos o funciones más complejas, al igual que ya vimos en los otros tipos de funciones.

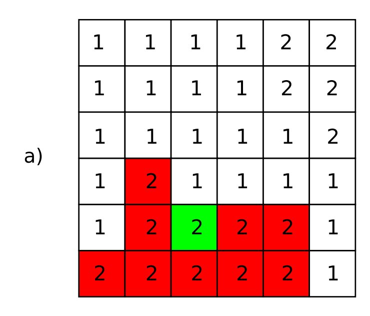
Los valores a asignar a cada clase pueden extraerse también de la propia capa de clases, no siendo necesaria otra capa. En este caso, estos valores resultantes suelen tener relación no con un parámetro adicional, sino con la geometría de cada clase. Por ejemplo, la superficie o el perímetro de cada tesela pueden recogerse como valores asociados a esta.

Este es un tipo de análisis muy frecuente en el estudio del paisaje, y el número de parámetros que pueden obtenerse por análisis zonal a partir de una única capa de clases es muy elevado. Junto a parámetros sencillos como la citada superficie o el perímetro, otros parámetros más complejos pueden servir para recoger la configuración estructural de las teselas, su riqueza y variabilidad, la fragmentación, etc [152].

15.2.4. Funciones globales

Las funciones globales son aquellas que utilizan la totalidad de valores de la capa para la obtención del resultado (Figura 15.16). Por su forma de operar, no generan exclusivamente nuevas capas como las anteriores funciones, sino tanto valores concretos como objetos geográficos de diversa índole.

Por ejemplo, los valores máximo y mínimo de la capa que son necesarios para normalizar esta se



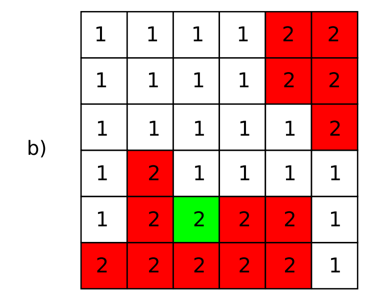


Figura 15.15: Formas de definir las clases para el análisis zonal. En verde, celda de análisis. En rojo, celdas a considerar en su misma clase para ser empleadas en el cálculo. a) asignación por valor con contigüidad, b) asignación únicamente por valor.

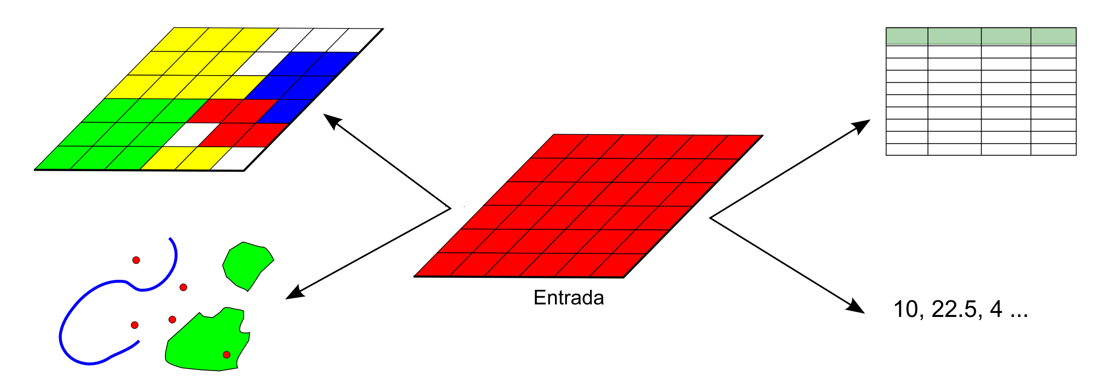


Figura 15.16: Las funciones de análisis global analizan el conjunto de valores de una capa para obtener un valor resultante, que puede ser tanto un objeto geográfico (capa ráster o vectorial) como un valor escalar sencillo, una tabla u otro tipo de resultado.

obtienen mediante una función global. Asimismo, el cálculo de un perfil entre dos puntos (Figura 15.17) o el trazado de una ruta óptima sobre una superficie de coste acumulado (que veremos en el apartado 20.4) son ejemplos de funciones globales que generan un resultado distinto de un mero valor numérico.

En el apartado dedicado a las funciones locales veíamos cómo la aplicación del operador *distinto de* entre dos capas de uso de suelo correspondientes a distintas fechas nos servía para localizar las zonas que habían experimentado cambios en su uso de suelo. Tras esta operación, la capa resultante contendrá un valor *verdadero*, habitualmente representado con un 1, en las zonas donde se da esa variación, y *falso*, codificado con 0, en las restantes. Si queremos cuantificar esa variación, podemos aplicar un operador global que sencillamente sume los valores de todas las celdas de la capa, lo cual dará como resultado el número total de celdas cuyo uso de suelo ha variado en el periodo de tiempo comprendido entre las dos fechas representadas por las capas de entrada.

Puesto que los operadores globales operan sobre la totalidad de la capa, a veces resulta conveniente o incluso necesario «eliminar» de esta los valores que no son de interés para el cálculo.

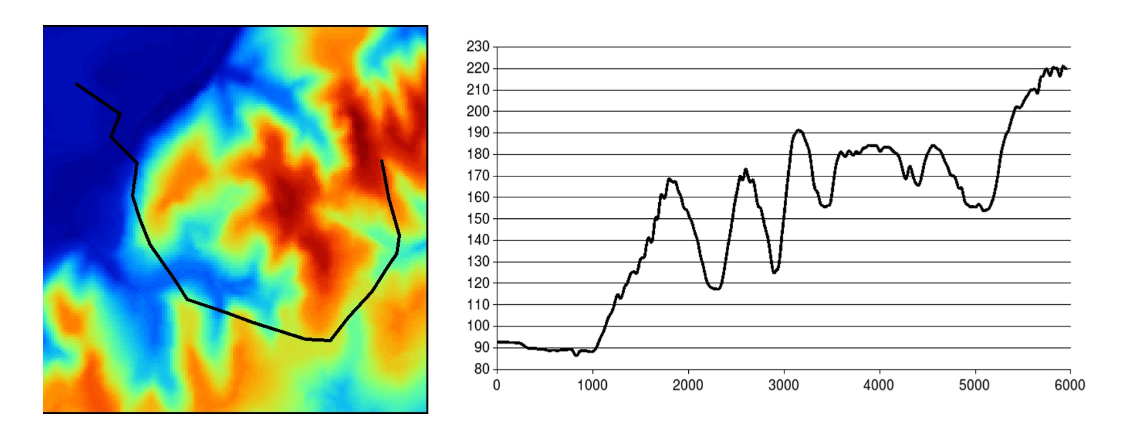


Figura 15.17: Dada una ruta y una capa, podemos obtener un perfil de valores a lo largo de la ruta mediante una función global.

No debemos olvidar que una capa ráster tiene una forma rectangular, la cual raramente se va a corresponder con la de la región de análisis, ya sea esta definida por un limite natural o no. El uso de máscaras que vimos en 15.2.1 es muy práctico a estos efectos.

Considérese, por ejemplo, que una curva hipsográfica que representa la distribución de alturas dentro de un área dada (habitualmente una unidad hidrológica), no tiene mucho sentido si se aplica a una región delimitada de forma «artificial» por los límites rectangulares de la capa. Resulta más lógico aplicar una máscara sobre la capa a analizar, de modo que la función global ignore las celdas que, aun estando en la capa, no están en la unidad de interés. Estas celdas tendrán asociado un valor de sin datos tras la aplicación de dicha máscara.

15.3. Preparación de capas

Las variables que manejamos en el álgebra de mapas son, como hemos visto en los ejemplos precedentes, capas ráster y valores escalares que podemos combinar con los anteriores. Para algunas de las funciones resulta necesaria únicamente una capa, mientras que para otras son necesarias varias.

En los ejemplos que hemos visto de combinación de varias capas, hemos dado siempre por supuesto que todas ellas tienen una estructura común. Es decir, que cubren una misma porción de terreno y lo hacen mediante una malla de celdas de las mismas dimensiones, con un mismo tamaño de celda y una misma georreferenciación. De este modo, un punto del terreno con coordenadas dadas queda reflejado en todas las capas en la misma celda i, j, y podemos operar con sus valores directamente para obtener un resultado correspondiente a dicho emplazamiento.

No obstante, a la hora de combinar capas es muy frecuente que estas tengan procedencias distintas y esta circunstancia no se dé. En tal caso, hay que preparar las capas para adecuarlas a un mismo marco geográfico sobre el que aplicar las funciones del álgebra de mapas de forma adecuada. Si este marco consiste en una malla de celdas de dimensiones $n \times m$, y las coordenadas de cada celda i, j son respectivamente x_{ij} e y_{ij}, deben calcularse los valores de las capas en esas coordenadas a partir de los valores en los marcos de

referencia originales. Este proceso se denomina *remuestreo.*

El remuestreo en realidad es una interpolación similar a la que veíamos en el capítulo 14, con la diferencia de que en este caso los puntos con datos no están distribuidos irregularmente sino de forma regular en una malla, con lo que podemos dar una expresión para la función interpolante en función de las celdas de origen situadas en torno a la coordenada en la que queremos calcular el nuevo valor (la del centro de cada celda en la capa remuestreada).

Los métodos más habituales de remuestreo son los siguientes:

- **Por vecindad**. Como ya vimos, no se trata en realidad de una interpolación como tal, pues simplemente crea la nueva malla situando nuevas celdas cuyos valores se calculan por mera vecindad, tomando el de la celda más cercana.
- **Bilineal**. Para una celda (i', j') en la nueva malla interpolada, su valor en función de los de las 4 celdas más cercanas a la misma en la malla original viene dado por la expresión.

$$\begin{aligned} z_{(i',j')} &= z_{(i,j)} R(-a) R(b) \\ &+ z_{(i,j+1)} R(a) R(-(1-b)) \\ &+ z_{(i+1,j)} R(1-a) R(b) \\ &+ z_{(i+1,j+1)} R(1-a) R(-(1-b)) \end{aligned} \tag{15.3.1}$$

donde $R(x)$ es una función triangular de la forma

$$R(x) = \begin{cases} x+1 & \text{si } -1 \leq x \leq 0 \\ 1-x & \text{si } -0 \leq x \leq 1 \end{cases} \tag{15.3.2}$$

- **Interpolación bicúbica**. La interpolación bicúbica es un método de interpolación multivariante bidimensional que emplea un polinomio de tercer grado para cada una de las direcciones. Son necesarias 16 celdas en lugar de las 4 de la bilineal, lo que hace que el método sea más exigente en términos de proceso. Para el caso habitual de emplear como función interpolante un spline cúbico, se tiene

$$\begin{aligned} z_{(i',j')} &= \sum_{m=-1}^{2} \sum_{n=-1}^{2} (z_{(i+m,j+n)} R(m-a) \\ &\cdot R(-(m-b))) \end{aligned} \tag{15.3.3}$$

$$\begin{aligned} R(x) &= \frac{1}{6}((x+2)_+^3 - 4(x+1)_+^3 \\ &+6(x)_+^3 - 4(x-1)_+^3) \end{aligned} \tag{15.3.4}$$

siendo

$$(x)_+^m = \begin{cases} x^m & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases} \tag{15.3.5}$$

Los métodos de remuestreo son un área muy desarrollada en el tratamiento de imágenes digitales, aunque, en la práctica, la mayoría de algoritmos existentes no presentan una diferencia notable con los anteriores (excepto con el remuestreo por vecindad) a la hora de aplicarlos sobre capas ráster de variables continuas en lugar de imágenes. Es por ello que su implementación y uso no es habitual en el caso de los SIG. En [153] puede encontrarse una buena introducción a otro tipo de funciones utilizadas para el remuestreo de imágenes.

A la hora de elegir uno u otro de los métodos anteriores, debe tenerse en cuenta, fundamentalmente, el tipo de información que contenga la capa. Una diferencia fundamental que debe tenerse siempre presente es que, de entre los métodos anteriores, el de vecino más cercano es el único que garantiza que los valores resultante existen como tales en la capa origen. Ello hace que este sea el único método que puede utilizarse a la hora de remuestrear capas de información categórica. Podemos ver claramente esto en la figura 15.18.

a)

1	3	5	9	6	0
3	5	2	2	3	4
7	2	8	1	1	4
2	5	6	6	5	3
1	9	3	4	2	8
5	7	2	1	8	6

b)

3	2	3
2	6	5
5	2	8

c)

3	2	3
2	5,56	5
5	2	8

Figura 15.18: El remuestreo de capas categóricas solo puede llevarse a cabo por vecindad. a) capa original, b) remuestreo por vecindad, c) remuestreo mediante splines (incorrecto, con valores incoherentes)

Se aprecia que en la capa remuestreada mediante interpolación bicúbica aparece un valor no entero producto de las operaciones matemáticas aplicadas, frente a los valores enteros que representan las categorías en la capa original. Los valores no enteros carecen de sentido, y hacen así que la capa remuestreada no sea válida.

Incluso si no apareciesen valores decimales, el remuestreo de capas categóricas por métodos distintos del vecino más cercano es conceptualmente incorrecto, ya que la realización de operaciones aritméticas con valores arbitrariamente asignados a las distintas categorías carece por completo de sentido.

Cuando se trabaje con imágenes directamente, es de interés el considerar esta misma circunstancia referente a los métodos de remuestreo aplicables en relación con la interpretación de la imagen que vaya a llevarse a cabo. La aplicación del remuestreo por vecindad es en la mayoría de los casos la opción a elegir, en especial cuando se va a proceder a un análisis de la imagen con posterioridad.

Aun en los casos de variables no categóricas, y aunque la elección del método de remuestreo no conduce de por sí a un resultado necesariamente erróneo, el proceso de remuestreo como tal sí que puede hacerlo si no se razona en función de la información contenida en la capa. Podemos ver esto claramente en el ejemplo de la figura 15.19.

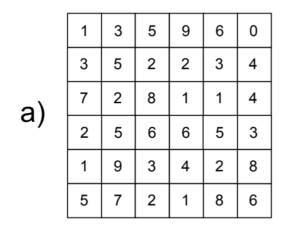
La capa original contiene información sobre el número de individuos de una especie que han sido encontrados en cada celda, de tal modo que representa la densidad de dicha especie. Si se modifica el tamaño de celda para hacerlo el doble de grande, la nueva celda tras el remuestreo cubre cuatro celdas de la capa original. Mientras que el remuestreo asignará a esa celda un valor promedio de las cuatro originales que engloba, el numero de individuos en ella será realmente la suma de ellos. Debe aplicarse un factor de reescala que relacione el área de la celda antes del remuestreo con el tamaño después del mismo, para así mantener la coherencia en el significado de la variable.

15.4. Extensión del álgebra de mapas

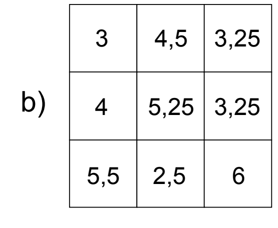
Aunque en la práctica los conceptos definidos por Tomlin son la base para la implementación genérica de algoritmos, diversos autores han intentado extender estos conceptos y formalizarlos de una forma más general. Aunque tratar estos sistemas escapa al alcance de este texto, resulta de interés mencionar algunas de las propuestas.

[154] propone un nuevo álgebra de mapas al que se incorporan predicados topológicos y direccionales. Esta definición permite la realización de operaciones que el álgebra de Tomlin no

a)



b)



c)

12	18	13
16	21	13
22	10	24

Figura 15.19: Dependiendo de la variable recogida en la capa, el proceso de remuestreo puede requerir operaciones adicionales para obtener un resultado correcto. a) capa con valores de conteos (número de individuos) por celda para una especie dada, b) capa tras remuestreo (incorrecta), c) capa tras remuestreo y aplicación de factor de reescala (correcta)

contempla, ya que, como demuestran, es un caso particular del anterior.

La propuesta de [155] con su *geo–álgebra* es distinta, y se encamina a una formalización matemática completa de las operaciones espaciales. En ella, no solo se contemplan los datos espaciales, sino también los procesos existentes. Así, se extiende no solo el álgebra de operaciones, sino el concepto de mapa a través de los nuevos conceptos de mapa *relacional* y *meta–relacional*. La integración de modelos basados, por ejemplo, en autómatas celulares, es posible dentro del marco de este geo–álgebra

Por último, y aunque no relacionada directamente con la información geográfica, el *álgebra de imágenes* definida por [156] guarda una gran similitud debida a la parecida naturaleza de los datos ráster y las imágenes como ya hemos comentado. Este álgebra de imágenes pretende establecer una notación algebraica con la cual expresar los algoritmos del procesado de imágenes, muchos de los cuales comparten una base conceptual común con los empleados en el análisis geográfico, y que veremos en el capítulo 17.

15.5. Resumen

El álgebra de mapas nos proporciona las herramientas necesarias para analizar capas ráster y obtener de ellas resultados derivados. Bien sea a partir de una capa, de dos, o de una batería de ellas, las funciones del álgebra de mapas definen un marco formal de procesos dentro del cual desarrollar los más diversos análisis. Distinguimos cuatro tipos básicos de funciones: locales, focales, zonales y globales.

De cada uno de ellos veremos numerosos ejemplos de aquí en adelante, ya que constituyen la base conceptual sobre la que se construyen la práctica totalidad de algoritmos de análisis de capas ráster. Estas funciones han de ir unidas a un manejo adecuado de las variables de entrada (las capas ráster), así como a una serie operadores que se aplican sobre las celdas que cada función define como objeto de análisis.

Geomorfometría y análisis del terreno

16.1. Introducción

Si es cierto que una inmensa parte de la información que manejamos es susceptible de ser georreferenciada, y ello justifica la necesidad de los SIG como herramientas, no es menos cierto que, de los procesos que estudiamos en un SIG, la gran mayoría requieren un conocimiento exhaustivo de la superficie sobre la que se desarrollan. Puesto que las actividades humanas se asientan en su práctica totalidad sobre la superficie terrestre, estudiar cualquier fenómeno asociado a ellas sin tener en cuenta esta implica dejar fuera una gran cantidad de información sumamente valiosa.

Este hecho concede por sí solo importancia a este capítulo, pues todas las disciplinas hacen uso de uno u otro modo de los resultados que en él veremos. Asimismo, justifica el hecho de que el Modelo Digital de Elevaciones (MDE) sea, con toda probabilidad, la capa de mayor importancia de cuantas podemos incorporar a cualquier análisis, ya que no solo su ámbito de utilidad es amplísimo, sino que, como veremos, el número de resultados que podemos obtener de su análisis es igualmente variado y abundante. De este modo, su versatilidad es tal que en prácticamente todas las situaciones será de interés el incorporar a un proceso de análisis bien el MDE directamente o bien alguna de las múltiples capas derivadas de este.

El estudio del terreno y la parametrización de sus características es, por tanto, pieza fundamental del análisis geográfico. La disciplina encargada de llevar a cabo este análisis es la *geomorfometría*, la cual se define como la «ciencia del análisis cuantitativo del relieve» [157, 158, 159]. La geomorfometría es una ciencia multidisciplinar con base en la geología, la geomorfología, la topografía, la matemática y otras diversas ramas de conocimiento. Esto no es de extrañar considerando lo anteriormente mencionado, ya que son muchas las disciplinas que han dedicado sus esfuerzos al análisis del relieve y el estudio de sus implicaciones sobre los distintos procesos.

Puesto que nos encontramos dentro del marco de los Sistemas de Información Geográfica, centraremos el contenido del capítulo en los aspectos de la geomorfometría que atañen a la utilización del MDE como fuente de información principal, así como en las formulaciones de análisis con base informática.

Por su especial relevancia, dedicaremos la primera parte del capítulo a tratar en profundidad los MDE como capas de datos particulares. Detallaremos tanto la creación del MDE como su preparación y tratamiento, con el objetivo de disponer de una capa de calidad acorde con las múltiples operaciones que sobre ella vamos a realizar.

Resumiendo el conjunto de análisis geomorfométricos, la aplicación de estos puede estructurarse en las siguientes etapas [72]:

1. Toma de datos de elevación en puntos establecidos (muestreo).
2. Generación de un modelo de la superficie a partir de los anteriores.
3. Corrección de errores y eliminación de elementos erróneos.
4. Cálculo de parámetros y elementos derivados.
5. Empleo de los parámetros y elementos anteriores.

Los dos primeros puntos no son estrictamente necesarios, puesto que ya hemos visto que puede obtenerse directamente un MDE a partir de tecnologías como el LiDAR, que hace innecesario tanto la toma de datos puntuales como la interpolación en base a estos. En caso de que esta última se lleve a cabo, ya vimos en el capítulo 14 las distintas técnicas existentes, aunque en este ampliaremos esa información para tratar las consideraciones específicas que deben tenerse en cuenta cuando la variable interpolada es la elevación.

16.2. El Modelo Digital de Elevaciones

El Modelo Digital de Elevaciones es la pieza clave del análisis geomorfométrico. Es el equivalente informatizado de la cartografía clásica de elevaciones tradicionalmente representada mediante curvas de nivel.

El MDE supone un enorme salto cualitativo frente a sus predecesores, ya que la utilización de un modelo numérico en lugar de uno gráfico da lugar a una diferencia conceptual en el ámbito del análisis superior a la que tiene lugar con otras capas de información. El MDE se convierte así en una fuente de información de primer orden para el estudio del medio.

Conviene, antes de proseguir, aclarar algunos apartados en relación con la terminología empleada, para evitar equívocos y confusiones. En relación con la cartografía digital de elevaciones, estos vienen motivadas por el empleo de dos términos distintos: Modelo Digital de Elevaciones (MDE) y Modelo Digital del Terreno (MDT).

Dentro de este texto consideraremos que un MDT es una «estructura numérica de datos que representa la distribución espacial de una variable cuantitativa y continua» [160], mientras que un MDE es un caso particular dentro de los MDT en el cual la variable recogida es la elevación.

El concepto de MDT, como vemos, es más genérico y no ha de guardar relación alguna con el relieve, puesto que, atendiendo a la definición anterior, podemos recoger variables tales como temperatura o la humedad en un punto dado, y hacerlo mediante un MDT. Tendríamos, por ejemplo, un Modelo Digital de Temperaturas. Expresado de otro modo, el MDT no es sino una capa con información continua como por ejemplo la relativa a elevaciones, aunque pudiendo recoger cualquier otra variable con esas características.

El empleo de la palabra *terreno* da lugar a que, con frecuencia, MDT y MDE sean empleados erróneamente como sinónimos en el uso diario de capas de elevaciones. En otros casos, por el contrario, sí se especifica una diferencia entre ambos conceptos, aunque esta no coincide con lo anterior. Algunos autores consideran que tanto el MDT como el MDE tienen relación con la elevación pero, mientras que el MDT recoge la elevación máxima en cada punto, incluyendo los elementos presentes tales como edificios, arboles, etc., el MDE recoge la información de la superficie sin tener esto en cuenta, en algo que podríamos denominar un *Modelo Digital de la Superficie*. Esta interpretación es más habitual en

lengua inglesa, aunque las acepciones también varían e incluso llegan a ser completamente contrarias, asignando al MDE las características del MDT y viceversa [161].

Aunque no han de ser necesariamente recogidos en capas ráster, los MDE generalmente se entienden por definición como entidades ráster e incluso algunos autores así lo reflejan explícitamente. Para [134], por ejemplo, un MDE es «una representación matricial regular de la variación continua del relieve en el espacio». Al igual que ya vimos en el capítulo 15 para el caso del álgebra de mapas genérico, la estructural regular de una malla ráster favorece la implementación de algoritmos de forma eficiente y sumamente más sencilla. Los algoritmos de análisis geomorfométrico no son una excepción a esto, y las formulaciones para capas ráster son mucho más abundantes que las basadas en otros modelos de almacenamiento, existentes de igual modo.

Junto a la información desarrollada en el capítulo 5 acerca de los distintos modelos de representación espacial, podemos ampliar esta y ver cómo dichos modelos se adaptan de modo particular a la representación del relieve.

Los formatos vectoriales tales como la representación mediante curvas de nivel son menos adecuados, ya que, como se dijo, no son óptimos para recoger variables de tipo continuo. Aunque interesantes para representar visualmente superficies continuas, no lo son tanto para su análisis, y las formulaciones existentes a este respecto dentro del análisis geomorfométrico son pocas.

Las redes de triángulos irregulares (TIN) son una forma mejor de recoger el relieve. No obstante, la implementación de algoritmos de análisis basados es TIN es más compleja y, por ello, menos frecuente. Los paquetes de software habituales no incluyen estas, por lo que en la práctica el análisis de un MDE en formato TIN es difícil de llevar a cabo.

El modelo más adecuado, tanto por sus características como por la facilidad con la que permite la implementación de algoritmos, es el modelo ráster, y será en este en el que se basen los análisis siguientes.

Las características de una capa ráster que definen la aplicabilidad de esta para un determinado análisis son sus resoluciones vertical y horizontal. En el caso del MDE, estos parámetros han sido ampliamente estudiados, así como su efecto en los resultados producidos a partir de dicho MDE. En [162] y [163] puede encontrarse más información al respecto

16.3. Creación y preparación del MDE

En muchas ocasiones, la información de elevaciones no se encuentra disponible como una capa ráster. Puesto que la mayoría de las herramientas de análisis se basan en este formato y el número de formulaciones aplicables es mucho mayor, se asume que el trabajo con un MDE dentro de un SIG con el objeto de llevar a cabo un análisis pormenorizado de este requiere disponer de una capa ráster de elevaciones.

Esta capa puede obtenerse a partir de la forma original en la que se encuentren los datos de elevación, haciendo uso de las diversas metodologías tales como los métodos de interpolación ya vistos.

Tanto si generamos la capa ráster como si la obtenemos directamente ya en este formato, es habitual que no se encuentre en las condiciones idóneas para ser sometida a todos los análisis que veremos en este capítulo. Será necesario trabajar sobre ella y prepararla para que sirva como una fuente de información fiable a la hora de obtener nuevos descriptores del terreno.

16.3.1. Creación del MDE

La creación del MDE implica en la mayoría de casos la utilización de métodos de interpolación. Un caso muy habitual es la creación a partir de curvas de nivel, con cuyos valores puede obtenerse una capa continua de elevaciones. También puede obtenerse a partir de datos puntuales tomados en campo, caso que es frecuente para estudios donde es necesaria una gran precisión, y en los que la cartografía existente no es suficiente, debiendo elaborarse esta.

La naturaleza de la información que se recoge en el MDE y el análisis a efectuar posteriormente con él hacen interesantes los siguientes comentarios sobre los distintos métodos de interpolación.

- **Distancia inversa**. El método de distancia inversa no es adecuado para la interpolación de MDE, pues genera abundantes depresiones de carácter artificial (ver 16.6.2) y elementos no naturales en el relieve.
- **Kriging**. Pese a ser un interpolador de calidad para el caso de datos de elevación, las superficies generadas son excesivamente suaves. Aunque la precisión del método es alta en términos cuantitativos, no refleja con precisión la configuración del relieve y sus accidentes, lo cual se hará notar en los análisis posteriores sobre dicho MDE.
  Otro problema del kriging es su excesiva sensibilidad a valores extremos estadísticamente diferentes del resto, ya que producen un número elevado de elementos morfológicos erróneos [72], dificultando la aplicación posterior de ciertos algoritmos.
- **Splines**. Los splines representan una de las mejores alternativas para la creación de MDE. Frente a las restantes metodologías, autores como [137] recomiendan el uso de splines para la obtención de MDE idóneos para el análisis.
- **Ajuste de funciones**. Como métodos globales, ajustar una función de tipo polinómico a los puntos de datos y asemejar el relieve a esta no es adecuado, ya que el relieve es altamente más complejo . El uso de funciones de grado superior y mayor complejidad, por otra parte, hace que puedan aparecer valores extremos irreales que desvirtúen por completo la superficie interpolada (ver 14.2.5). Aun así, veremos más adelante que este tipo de ajustes se utilizan no para la creación del MDE sino para el análisis de las propiedades de este, ajustando localmente funciones con los valores de un entorno definido de celdas.

Las anteriores son características generales de los métodos que han de considerarse de una u otra forma según sea el terreno a interpolar. Así, si el kriging adolece de una excesiva suavidad en la forma de la superficie calculada, no será un buen método para zonas montañosas con relieves escarpados. Si conocemos, sin embargo, que el relieve de la zona a interpolar es suave y redondeado, las deficiencias del kriging no serán tales en este caso. Una vez más, el conocimiento de la variable interpolada es un elemento básico para la aplicación correcta de uno u otro método.

Es de interés reseñar que los problemas en el MDE que hacen necesaria la preparación de este antes de su análisis vienen derivados de deficiencias no cuantitativas sino especialmente cualitativas, donde las formas del relieve recogidas no se corresponden con las reales. Por ello, es importante que el MDE generado refleje correctamente la conformación de la topografía, además de disminuir el error cuantitativo, puesto que la precisión cualitativa es en muchos casos de mayor importancia para obtener con posterioridad resultados correctos [164]. En este sentido, existen métodos particulares específicamente diseñados para la creación de MDE, cuyos resultados son más adecuados para un análisis directo y no requieren parte de la preparación que se hace necesaria al aplicar otros métodos de interpolación. Esto es

así debido a que los fundamentos de esta se incorporan ya en el proceso de interpolación y cálculo de nuevos valores.

El más destacable de estos métodos es ANUDEM [165], que toma como partida datos en forma de curvas de nivel. Además de utilizar los valores de elevación para interpolar, analiza la forma de las curvas de nivel para derivar otros resultados que sirven de apoyo. Asimismo, puede aportarse información adicional tal como la red de drenaje existente, en caso de que se disponga de ella, o los límites de barrancos u otras zonas donde el relieve no presente continuidad. Todo ello garantiza el buen comportamiento del MDE en el análisis posterior y su coherencia hidrológica.

16.3.2. Preparación del MDE

La preparación del MDE hace referencia al filtrado de este para eliminar las posibles deficiencias que contiene, y muy especialmente a los elementos relacionados con el análisis hidrológico. En el primer caso, se trata de eliminar información incorrecta presente en el MDE, mientras que en el segundo caso se trata de trabajar con elementos de este que dificultan la aplicación de ciertas formulaciones y algoritmos. Estos elementos no tienen que ser necesariamente erróneos

En términos generales, los objetivos que se persiguen con la preparación del MDE son:

- Eliminación de errores groseros.
- Descripción más precisa del relieve.
- Descripción más precisa de los procesos hidrológicos y ecológicos que tienen lugar.

La corrección de los valores del MDE enfocada a estos dos primeros objetivos incluye muchos factores distintos que han de tenerse en cuenta. Entre ellos, cabe mencionar.

- **Errores incorporados por las distintas fuentes de datos** o el instrumental empleado para la recogida de los mismos, así como por las metodologías empleadas en el proceso de estos (por ejemplo, las distintas técnicas de interpolación). Pueden corregirse conociendo las deficiencias de cada una de ellas, que dan información sobre el tipo de errores que pueden aparecer
- **Presencia de valores discrepantes muy improbables** (*outliers*). Estos pueden eliminarse mediante técnicas estadísticas, en función de los valores de las celdas vecinas [166]. La aplicación de métodos geoestadísticos como el kriging permite igualmente la estimación del valor esperado en una celda a partir de los valores de las circundantes.
- **Presencia de zonas sin datos**. Especialmente relevante en el caso de MDE procedentes de teledetección, ya que pueden existir zonas vacías por errores de captura, o bien por ser combinación de varias imágenes entre las que existen áreas no cubiertas. Estas zonas vacías pueden rellenarse mediante métodos de interpolación, siempre que estas no impliquen extrapolación y excedan de un tamaño máximo lógico.

Para una descripción más detallada de estas metodologías puede consultarse [72].

Además de la corrección de valores, la mayor parte de la preparación del MDE, imprescindible en muchos casos, es sin embargo de tipo hidrológico. Esta debe llevarse a cabo cuando se vayan a analizar parámetros que tengan relación con el comportamiento hidrológico de las celdas, y más concretamente todas aquellas que empleen algún algoritmo de asignación de direcciones de flujo de los que más adelante veremos.

La preparación del MDE corrige este comportamiento para solventar los dos principales problemas que se presentan a la hora de asignar direcciones de flujo: la presencia de zonas llanas y la existencia de depresiones. Ambos elementos hacen imposible el análisis hidrológico en aque-

llas zonas en las que aparecen, siendo necesario corregirlas previamente.

Para una mejor comprensión de las formulaciones que permiten corregir la presencia de zonas llanas o depresiones, es necesario conocer con anterioridad los algoritmos de dirección de flujo, ya que son estos los afectados por ellas. Por esta razón, veremos las soluciones a estos problemas en una sección posterior de este capítulo.

16.4. Modelos matemáticos locales del MDE

Gran parte de los análisis que vamos a realizar a partir del MDE tienen una base matemática. Conceptos tales como la pendiente o la orientación de una superficie en un punto dado pertenecen al ámbito de la geometría diferencial, y la aplicación de estos a la superficie recogida en un MDE requiere una caracterización matemática del mismo.

Sea una función que represente a nuestro Modelo Digital de Elevaciones, de la forma.

$$z = f(x, y) \tag{16.4.1}$$

donde x e y representan las coordenadas respectivas en los ejes x e y del plano horizontal, y z la elevación en el punto definido por dichas coordenadas. De existir una función como esta y poder disponerse de una expresión explícita de ella, contendría en sí toda información del MDE. Sin embargo, parece lógico pensar que, como ya se mencionó al tratar el ajuste de funciones como método de interpolación, las formas naturales del relieve no son definibles con dicha sencillez dada la complejidad de este.

Para poder dar una caracterización matemática al MDE debemos hacerlo localmente, utilizando una función distinta para cada celda. Acudiendo a los conceptos del álgebra de mapas, emplearemos una función de análisis focal que, dada una celda, toma un entorno reducido de esta y describe dicho entorno con una expresión matemática ajustada a este. La extensión limitada de la ventana de análisis hace que no sea irreal el pretender ajustar dicha función, como sí lo era en el caso de hacerlo para todo el MDE.

Las funciones más habituales son las de tipo cuadrático, que garantizan en mayor medida la corrección de los resultados derivados, tales como pendientes [167] o curvaturas [168] (veremos en breve qué significan estos parámetros y cómo se calculan).

$$\begin{aligned} f(x,y) &= ax^2y^2 + bx^2y + cxy^2 \\ &\quad + dx^2 + ey^2 \\ &\quad + fxy + gx + hy + k \end{aligned} \tag{16.4.2}$$

[15] propone el caso particular

$$z = \frac{rx^2}{2} + sxy + \frac{ty^2}{2} + px + qy + z_0 \tag{16.4.3}$$

Con seis parámetros que calcular, la ventana habitual 3×3, con nueve valores, es suficiente para llevar a cabo un ajuste por mínimos cuadrados. Empleando la notación introducida en 15.2.2 para dicha ventana, los valores de los parámetros quedan determinados de la siguiente forma:

$$\begin{aligned} p &= \frac{z_3 + z_6 + z_9 - z_1 - z_4 - z_7}{6\Delta s} \\ q &= \frac{z_1 + z_2 + z_3 - z_7 - z_8 - z_9}{6\Delta s} \\ r &= (z_1 + z_3 + z_4 + z_6 + z_7 + z_9 \\ &\quad -2(z_2 + z_5 + z_8))/3\Delta s^2 \end{aligned}$$

$$
\begin{aligned}
s &= \frac{-z_1 + z_3 + z_7 - z_9}{4\Delta s^2} \\
t &= (z_1 + z_2 + z_3 + z_7 + z_8 + z_9 \\
&\quad -2(z_4 + z_5 + z_6)/3\Delta s^2 \\
z_0 &= (5z_5 + 2(z_2 + z_4 + z_6 + z_8) \\
&\quad -(z_1 + z_3 + z_7 + z_9))/9
\end{aligned} \tag{16.4.4}
$$

siendo Δs el tamaño de celda.

Mientras que la superficie definida por esta función no ha de pasar necesariamente por ninguna de las celdas, [169] propone utilizar la expresión completa de 16.4.2, que expresaremos como sigue para que los cálculos posteriores resulten más sencillos:

$$
\begin{aligned}
z &= Ax^2y^2 + Bx^2y + Cxy^2 + \frac{rx^2}{2} \\
&\quad +sxy + \frac{ty^2}{2} + px + qy + D
\end{aligned} \tag{16.4.5}
$$

Esta función, ajustada mediante polinomios de Lagrange, sí garantiza que la superficie pase exactamente por los puntos de la ventana de análisis.

En este caso, los valores de los coeficientes son los siguientes:

$$
\begin{aligned}
p &= \frac{z_6 - z_4}{2\Delta s} \\
q &= \frac{z_2 - z_8}{2\Delta s} \\
r &= \frac{z_4 + z_6 - 2 \cdot z_5}{\Delta s^2} \\
s &= \frac{-z_1 + z_3 + z_7 - z_9}{4\Delta s^2} \\
t &= \frac{z_2 + z_8 - 2z_5}{\Delta s^2} \\
A &= ((z_1 + z_3 + z_7 + z_9) \\
&\quad -2(z_2 + z_4 + z_6 + z_8) + 4 \cdot z_5/4)\Delta s^4 \\
B &= \frac{(z_1 + z_3 - z_7 - z_9) - 2(z_2 - z_8)}{4\Delta s^3} \\
C &= \frac{(-z_1 + z_3 - z_7 + z_9) - 2(z_6 - z_4)}{4\Delta s^3} \\
D &= z_5
\end{aligned} \tag{16.4.6}
$$

Junto a las anteriores, probablemente las más habituales, encontramos otras como [170], idéntica a la de [15] salvo por el hecho de que obliga a la función a pasar por la celda central, o [171], así como funciones de orden superior o bien ajustes de planos en lugar de superficies cuadráticas [172]. Igualmente, puede generalizarse lo anterior al empleo de polinomios de grado n [151].

Independientemente del método empleado, disponemos de una caracterización matemática del relieve a nivel local, lo cual nos permite ya aplicar las herramientas matemáticas con las que obtener valores derivados.

16.5. Análisis morfométrico

El análisis morfométrico del MDE estudia el relieve y la orografía de la superficie definida por este, caracterizándolo a través de parámetros adicionales. Recurriendo una vez más a los conceptos del álgebra de mapas, estos parámetros se basan en su gran mayoría en un análisis focal. Este análisis emplea la caracterización matemática que vimos en el punto anterior, así como formulaciones estadísticas sencillas. En función de ello distinguimos dos grupos principales:

- Medidas geométricas
- Medidas estadísticas

Puesto que la mayoría de análisis son de tipo focal, se necesita definir una ventana de análisis. El tamaño de esta ventana depende del tipo de medida a calcular. En el caso de medidas estadísticas, esta puede escogerse libremente, así como su forma. En el caso de las medidas geométricas, estas requieren de un modelo matemático

del terreno, por lo que la ventana de análisis viene condicionada al modelo escogido. En realidad, una vez establecida la función que define localmente la superficie del MDE, hacemos uso directo de su expresión, motivo por el cual, según sea el modelo, se usarán unos u otros valores de los pertenecientes a la ventana de análisis.

16.5.1. Medidas geométricas

Las medidas geométricas derivan de la caracterización matemática local del relieve, sobre la cual se aplican conceptos fundamentales de geometría diferencial. Estas medidas describen las propiedades geométricas de la porción de relieve contenida en la ventana de análisis.

Existen formulaciones no basadas en un análisis matemático del terreno [173], aunque nos centraremos en aquellas con una base matemática, por su mayor idoneidad. Estas formulaciones que no hacen uso de la caracterización matemática local del terreno permiten obtener valores de pendiente y orientación, pero no de otros parámetros tales como curvaturas.

Medidas basadas en derivadas de primer grado

Los parámetros geométricos más básicos parten del estudio de las derivadas parciales de primer orden de la superficie, y se obtienen a partir del *gradiente* de la superficie. El gradiente es un concepto básico del cálculo vectorial y se expresa en función de las antedichas primeras derivadas parciales, según

$$\nabla\overline{\mathbf{Z}} = \left(\frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}\right) \tag{16.5.1}$$

Dado un campo escalar tal como el representado por el MDE, el gradiente es un campo vectorial orientado en la dirección en la que el campo escalar experimenta una mayor variación.

Las dos propiedades principales que podemos obtener del gradiente son dos: su longitud o módulo y su dirección. Estas propiedades, considerando la interpretación geomorfométrica que le damos por calcularse a partir del MDE, constituyen dos de los parámetros más importantes que pueden extraerse de este: la pendiente y la orientación.

La pendiente refleja la variación máxima de la elevación y se define en función del gradiente como

$$s = \arctan(|\nabla\overline{\mathbf{Z}}|) \tag{16.5.2}$$

Este valor representa el ángulo entre el plano horizontal y el tangencial a la superficie en el punto considerado. Este último queda definido por el gradiente, que es normal al mismo. El horizontal es, a su vez, normal a la curva de nivel que pasa por el punto.

La pendiente obtenida de la expresión anterior viene expresada en unidades angulares tales como radianes o grados, aunque es habitual hacerlo en porcentaje. Para ello, utilizamos la expresión

$$s = \tan(s) \cdot 100 \tag{16.5.3}$$

Si utilizamos el modelo matemático de [15], derivando la expresión dada obtenemos:

$$\begin{aligned}\frac{\partial z}{\partial x} &\approx \frac{\partial(\frac{rx^2}{2} + sxy + \frac{ty^2}{2} + px + qy + z_0)}{\partial x} \\ &= rx + sy + p \end{aligned} \tag{16.5.4}$$

$$\begin{aligned}\frac{\partial z}{\partial y} &\approx \frac{\partial(\frac{rx^2}{2} + sxy + \frac{ty^2}{2} + px + qy + z_0)}{\partial y} \\ &= ty + sx + q \end{aligned} \tag{16.5.5}$$

El valor que buscamos es el de la celda central, que representa el origen de coordenadas. Por tanto, $x = y = 0$, y según 16.4.4 tenemos que

$$p = \frac{\partial z}{\partial x} = \frac{z_3 + z_6 + z_9 - z_1 - z_4 - z_7}{6 \cdot \Delta s} \quad (16.5.6)$$

$$q = \frac{\partial z}{\partial y} = \frac{z_1 + z_2 + z_3 - z_7 - z_8 - z_9}{6 \cdot \Delta s} \quad (16.5.7)$$

Con estos valores ya tenemos el vector gradiente, y aplicando la ecuación 16.5.2 obtenemos el valor de la pendiente en la celda.

$$s = \arctan(\sqrt{p^2 + q^2}) \quad (16.5.8)$$

Si empleamos en su lugar el modelo introducido en 16.4.5, derivando y haciendo $x = y = 0$, se obtiene también que

$$p = \frac{\partial z}{\partial x} \ ; \ q = \frac{\partial z}{\partial y} \quad (16.5.9)$$

Y según 16.4.6

$$\frac{\partial z}{\partial x} = \frac{z_6 - z_4}{2\Delta s} \quad (16.5.10)$$

$$\frac{\partial z}{\partial y} = \frac{z_2 - z_8}{2\Delta s} \quad (16.5.11)$$

Como ejemplo, una capa de pendientes resultante de utilizar esta última formulación puede verse en la figura 16.1.

Podemos expresar ambas formulaciones como el resultado de convoluciones (ver 15.2.2), mediante un núcleo. Por ejemplo, las parciales según la metodología de [15] se corresponden con los núcleos siguientes:

$\frac{1}{-6\,\Delta s}$	0	$\frac{1}{6\,\Delta s}$
$\frac{1}{-6\,\Delta s}$	0	$\frac{1}{6\,\Delta s}$
$\frac{1}{-6\,\Delta s}$	0	$\frac{1}{6\,\Delta s}$

$\frac{1}{-6\,\Delta s}$	$\frac{1}{-6\,\Delta s}$	$\frac{1}{-6\,\Delta s}$
0	0	0
$\frac{1}{6\,\Delta s}$	$\frac{1}{6\,\Delta s}$	$\frac{1}{6\,\Delta s}$

Como ya sabemos, la estructura ráster del MDE es similar a la de una imagen digital. Esto hace que no solo sea sencillo combinar el MDE con imágenes aéreas o satelitales, sino que también en las técnicas y algoritmos existe una notable similitud. En particular, veremos más adelante en el capítulo dedicado al análisis de imágenes que los núcleos anteriores se conocen como *núcleos de Prewitt* y son frecuentemente utilizados para la detección de bordes.

Muy relacionado con la pendiente, encontramos otro parámetro muy sencillo: la superficie real de una celda. Si tenemos una capa con un tamaño de celda Δs, el área ocupada por la misma será $A = (\Delta s)^2$. Este area es un área planimétrica, ya que en ningún momento consideramos que la celda tiene inclinación, y no se corresponde con, por ejemplo, el área real que se encuentra disponible para los seres vivos que habitan en esa celda.

Una manera sencilla de calcular el área real de una celda es empleando la siguiente fórmula [174]:

$$A' = \frac{A}{\cos(s)} \quad (16.5.12)$$

La simplicidad de la fórmula, sin embargo, hace que, por regla general, el área quede sobreestimada [175], al contrario que en el caso de utilizar directamente el área planimétrica. [176] propone como alternativa el utilizar el conjunto de triángulos tridimensionales que se establecen entre la celda central y cada una de las adyacentes, y sumar las áreas de todos ellos. De este modo, se obtiene una aproximación más realista.

Junto con la pendiente, la orientación es el otro parámetro básico que calculamos a partir del gradiente. Esta queda definida como [177]:

$$a = 180 - \arctan\left(\frac{q}{p}\right) + 90\frac{p}{|p|} \quad (16.5.13)$$

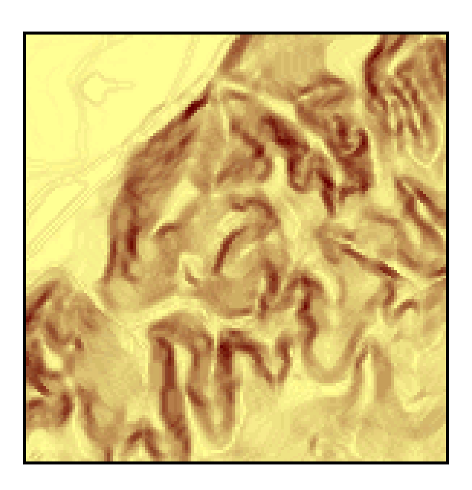


Figura 16.1: Capa de pendientes obtenida a partir del MDE



Figura 16.2: Capa de orientaciones obtenida a partir del MDE.

siendo p y q las derivadas parciales en las dos direcciones x e y, según la notación que venimos empleando.

El resultado viene expresado en grados sexagesimales. Se considera por convención que el Norte marca la orientación cero, aumentando los valores en sentido horario.

La formula anterior no se encuentra definida en aquellas celdas para las cuales $\partial z/\partial y = 0$. [178] propone utilizar en su lugar la siguiente fórmula:

$$a = -90 \cdot [1 - \mathrm{sign}(q)] \cdot [1 - |\mathrm{sign}(p)|] + 180 \cdot [1 + \mathrm{sign}(p)] - \frac{180}{\pi} \cdot \mathrm{sign}(p) \cdot \arccos \frac{-q}{\sqrt{p^2 + q^2}} \tag{16.5.14}$$

siendo $\mathrm{sign}(x)$ la función definida por

$$\mathrm{sign}(x) = \begin{cases} 1 & \text{si } x > 0 \\ 0 & \text{si } x = 0 \\ -1 & \text{si } x < 0 \end{cases} \tag{16.5.15}$$

En este caso, la orientación no está definida cuando la pendiente es nula.

La figura 16.2 muestra una capa con valores de orientación.

Es de reseñar que la orientación es un parámetro circular, ya que a y $a + 360°$ indican la misma dirección. Por esta razón, debe tenerse cuidado al realizar análisis estadísticos sobre los valores de orientación. Una posible solución que elimina este hecho es trabajar con las proyecciones $\cos(a)$ y $\sin(a)$ [179] (ver 13.3).

También a la hora de crear una representación gráfica de una capa de orientaciones debe tenerse en cuenta la naturaleza circular del parámetros. Nótese como en la figura 16.2 la escala de colores también es circular, de tal modo que el valor máximo (360) tiene asociado el mismo color que el mínimo (0), ya que en realidad representan una misma orientación.

Tanto la orientación como la pendiente pueden representarse de modo simultaneo no como una capa nueva de valores, sino mediante elementos vectoriales (líneas) que representen el gradiente, las cuales pueden acompañar a la

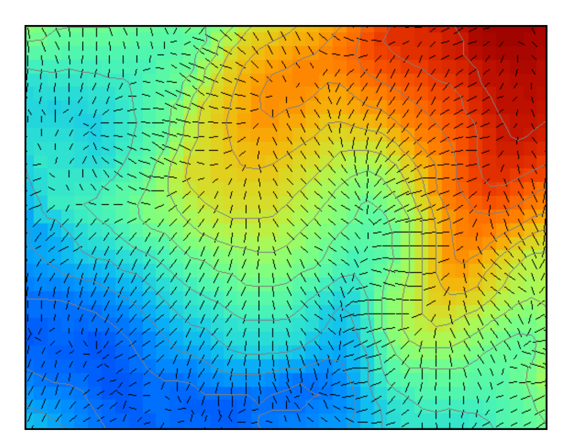


Figura 16.3: Representación del gradiente mediante elementos lineales. Para una mejor interpretación se han añadido curvas de nivel cada 10 metros.

propia capa de elevaciones a partir de la cual se calculan (Figura 16.3).

Aunque la pendiente y la elevación se aplican habitualmente a una capa de elevaciones como todos los restantes parámetros que vemos en este capítulo, no ha de limitarse su aplicación a estas. Para cualquier otra variable de la cual queramos ver cómo varía en el espacio, ambos parámetros son perfectamente aplicables.

Como veremos más en detalle en las próximas secciones, la orientación tiene relación directa con el comportamiento de los flujos sobre el terreno y sirve por tanto para el análisis hidrológico. Los análisis que veremos son de tipo regional, aunque la orientación puede usarse también en análisis focales para definir parámetros relacionados con el comportamiento hidrológico, tales como el *índice de convergencia* [180].

Otro de los usos habituales de los MDE es el relativo a la insolación. Tradicionalmente, la propia orientación se ha utilizado para evaluar si una zona se encuentra situada en solana o en umbría. Con la mayor potencia que el cálculo basado en un MDE nos brinda, podemos obtener resultados más precisos.

Una primera aproximación mediante un modelo sencillo se obtiene combinando los valores de orientación y pendiente en los denominados *relieves sombreados*. Estos mapas intentan reflejar la topografía del MDE de forma más realista, constituyendo además representaciones muy intuitivas. Por esta razón, se han empleado para resaltar otros mapas y enriquecerlos visualmente [171], tal y como puede verse en la figura 16.4. Empleando transparencias, puede reflejarse en una única imagen la información de una capa ráster y el relieve subyacente.

Para una posición dada del Sol, definida por su azimut ϕ y su elevación sobre el horizonte θ, la insolación sobre una celda de pendiente s y orientación a viene expresada por la ecuación [181]:

$$F = \frac{100\tan(s)}{\sqrt{1+\tan^2(s)}} \cdot \left[\frac{\sin(\theta)}{\tan(s)} - \cos(\theta)\sin(\phi - a)\right] \tag{16.5.16}$$

Los valores se sitúan en el rango 0–100. La capa resultante puede emplearse no solo para

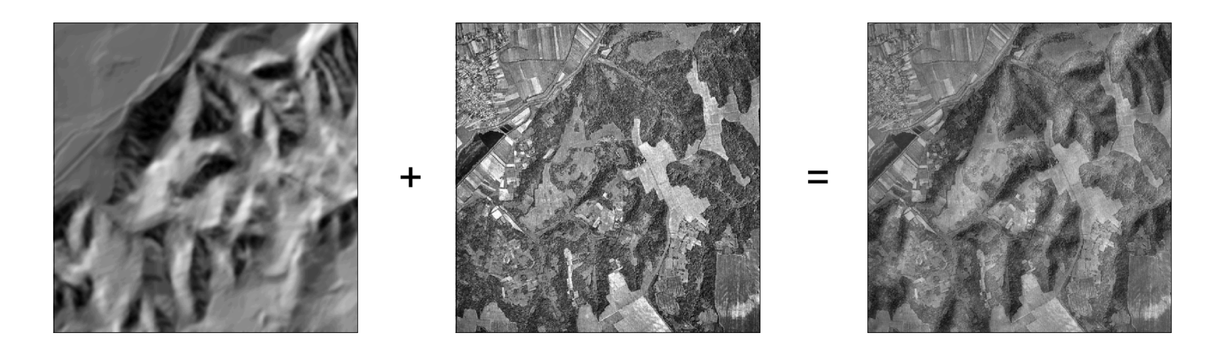


Figura 16.4: Utilización de una capa de relieve sombreado para mejorar la visualización de una imagen aérea.

mejorar visualmente otras, sino, por ejemplo, para corregir el efecto del relieve en imágenes de satélite. [182] [160].

El modelo de insolación anterior es muy simple y puede mejorarse notablemente. Por una parte, para reflejar realmente la insolación debemos considerar las sombras que pueden proyectar otros elementos del relieve situados en torno a la celda considerada. Es decir, el estudio riguroso de la insolación no puede ser un análisis focal. Por otra parte, los valores sin unidades en el rango 1–100 que son válidos para la mejora visual de otras capas, serían mucho más útiles para otras aplicaciones si contuvieran información en unidades reales de energía radiante. Y por último, esta energía radiante debería poder expresarse no soló instantánea para una posición dada del Sol, sino acumulada a lo largo de un periodo de tiempo tal como un día, una estación o un año completo. De este modo, sería mucho más útil para, por ejemplo, desarrollar modelos de desarrollo de especies vegetales o animales, que guardan una relación directa con la energía solar recibida.

Respecto al estudio de la insolación no como un parámetro local, sino teniendo en cuenta las sombras causadas por el relieve circundante, se aplican análisis de visibilidad para ver si existe conexión visual entre el Sol y una celda dada o, por el contrario, el relieve bloquea la línea de visión. En 16.7 se detallará lo necesario para llevar a cabo este análisis.

En la figura 16.5 puede verse esto gráficamente.

Para analizar la insolación a lo largo de un periodo dado, basta dividir este en intervalos y estudiar la posición del Sol en cada uno.de ellos. Los ángulos ϕ y θ se obtienen en cada instante a partir de modelos astronómicos sencillos. En [183] se encuentran detalladas las expresiones correspondientes. Del análisis de todos esos intervalos obtenemos el tiempo de insolación total como suma del de todos los intervalos en los que el Sol es visible desde la celda, así como un valor total de insolación como suma de las energías recibidas en ellos.

Para estimar dicha energía correspondiente a un intervalo, estudiamos las pérdidas de energía debidas a la transmisión desde el Sol a la Tierra, y las correspondientes al relieve.

Comenzando partir de la magnitud E recibida en el exterior de la atmósfera, se puede calcular la recibida por una superficie plana —es decir, horizontal— situada a una altura h sobre el nivel del mar, según los pasos siguientes.

Se tiene que

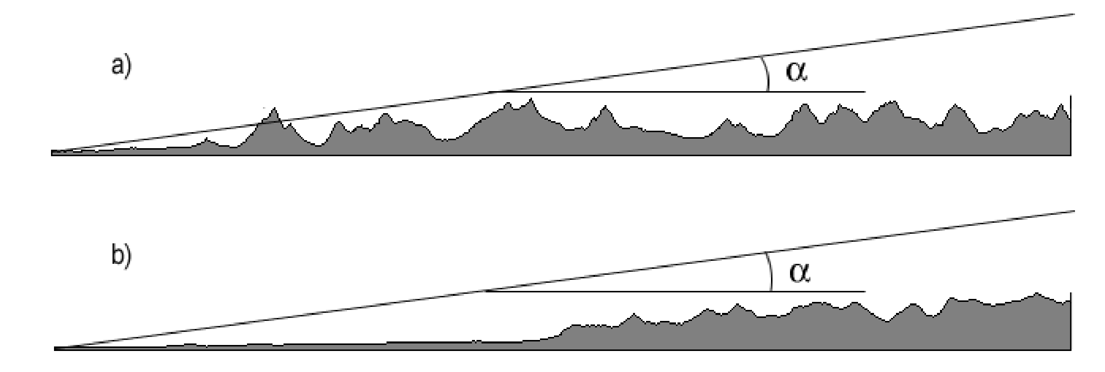


Figura 16.5: Zonas en sombra (a) frente a zonas bajo insolación (b), por comparación entre el perfil del terreno y la línea de visión entre la celda problema y el Sol.

$$E = E_0(1 + 0,034\cos(360d/365)) \quad (16.5.17)$$

donde E_0 es la conocida constante solar de valor 1367W m^{-2} y d el número de día del año.

El paso por la atmósfera reduce el valor de E según

$$E' = E\tau^{M_h} \quad (16.5.18)$$

donde τ es la transmisividad del aire, generalmente de valor $\tau = 0,6$, y M_h la masa del aire, la cual se evalúa de acuerdo con la expresión

$$M_h = M_0 C_p \quad (16.5.19)$$

siendo

$$M_0 = \sqrt{1229 + (614\sin\alpha^2) - 614\sin\alpha}$$
$$C_p = \left(\frac{228 - 0,0065h}{288}\right)^{5,256} \quad (16.5.20)$$

Por último, el relieve de la celda modifica la energía recibida según

$$E'' = E'\cos\alpha \quad (16.5.21)$$

El valor de $\cos\alpha$ se calcula a partir de las orientaciones y pendientes tanto del Sol como de la celda, haciendo uso de la expresión

$$\cos\alpha = \cos\theta\sin s + \cos(\phi_s - a)\sin\theta\cos s \quad (16.5.22)$$

No obstante, la introducción de factores como la masa de aire añaden una cierta subjetividad al modelo al ser datos estimados a su vez a partir de otros parámetros, pudiendo restar en cierta medida aplicabilidad a este planteamiento, según casos. Junto a esto, otra serie factores tales como el albedo debido a la nubosidad no son considerados en el anterior modelo, y serían difíciles de integrar. Y por último, la radiación difusa debe considerarse, pues su aporte es relevante [72]

[160] propone reflejar las diferencias que existen entre cada celda mediante la expresión de las energías de modo relativo, tomando una referencia válida para ello. En particular, es de interés dar las energías como porcentajes de la que sería recibida por una superficie horizontal, pues resulta sencillo y altamente operativo a efectos de calcular una malla de insolación como la que buscamos. Se define así el concepto de *índice de radiación*, según la expresión

$$I_{\mathrm{rad}} = \sum_{i=1}^{n} \frac{\cos\sigma}{\cos(\pi/2 - \gamma_s)} \tag{16.5.23}$$

Aplicando este concepto al cálculo de la energía en una celda dada, esta puede obtenerse para dicha celda, y sobre un intervalo de tiempo dividido en n periodos, mediante la expresión

$$E = E_0 \sum_{i=1}^{n} \frac{\cos\sigma}{\cos(\pi/2 - \gamma_s)} \tag{16.5.24}$$

E_0 es la energía recibida por una superficie horizontal.

Esta aproximación es sumamente útil si se dispone de datos puntuales procedentes de piranómetro, ya que, tomando dichos datos como referencia —esto es, conociendo el valor exacto de E_0—, se podrían calcular las restantes celdas de la malla aplicando la anterior relación.

Medidas basadas en derivadas de segundo grado

Las medidas geométricas basadas en derivadas de segundo grado aportan información sobre la concavidad o convexidad de la superficie en un punto dado. El parámetro que expresa esa información se denomina *curvatura*.

Al contrario que sucedía con la pendiente o la orientación, no existe una única curvatura, sino un número infinito de ellas. Para definir una curvatura particular necesitamos no solo el punto donde calcularla, sino también un plano que corte a la superficie en el mismo. Este plano ha de contener al vector normal a la superficie, y la intersección de dicho plano con esta da lugar a lo que se denomina una *sección normal* (Figura 16.6), sobre la cual se estudia la curvatura. La curvatura de la curva plana que representa la sección normal es el inverso del radio del circulo que se ajusta a la curva en el punto dado (Figura 16.7). Por tanto, tiene dimensiones $[\mathrm{L}^{-1}]$.

Matemáticamente, la curvatura k de la curva $z(x)$ viene dada por:

$$k = \frac{\frac{\mathrm{d}^2 z}{\mathrm{d}z^2}}{\left[1 + \left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2\right]^{1,5}} \tag{16.5.25}$$

De entre los planos que pueden tomarse, algunos aportan más información que otros. Los planos normales a la superficie que son paralelo y normal a la curva de nivel (mutuamente ortogonales) resultan de especial interés. Estos definen, respectivamente, las curvaturas denominadas horizontal y vertical.

Junto con la notación introducida en 16.5.9, utilizaremos la siguiente para simplificar las expresiones:

$$r = \frac{\partial^2 z}{\partial x^2}; \qquad s = \frac{\partial^2 z}{\partial x \partial y}; \qquad t = \frac{\partial^2 z}{\partial y^2} \tag{16.5.26}$$

Derivando tanto la fórmula de [15] como la de [169], y dadas la expresiones escogidas para estas, se tiene que las derivadas parciales r, s y t coinciden respectivamente con los coeficientes r, s y t de aquellas, calculados en la sección 16.4.

Empleando la notación de 16.5.26, la fórmula de la curvatura vertical es la siguiente [15, 184]:

$$kv = \frac{-p^2 \cdot r + 2pqrs + q^2 t}{(p^2 + q^2)\sqrt{(1 + p^2 + q^2)^3}} \tag{16.5.27}$$

Por convenio, las zonas convexas tienen curvatura negativa y las cóncavas curvatura positiva.

En el caso de valores convexos de curvatura vertical, estos indican un aumento en la pendiente según se sigue la línea de máxima inclinación. Puesto que esta es la línea que marca la dirección del agua al fluir, las zonas con curvatura vertical convexa se corresponden con aquellas en

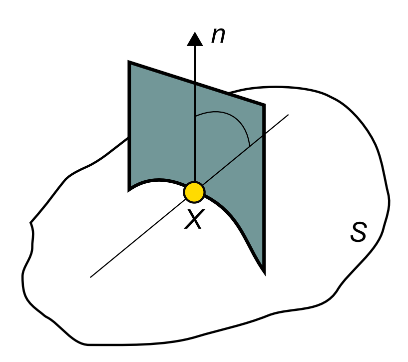


Figura 16.6: Una sección normal es la resultante de la intersección de la superficie con un plano que contiene al vector normal a esta en un punto.

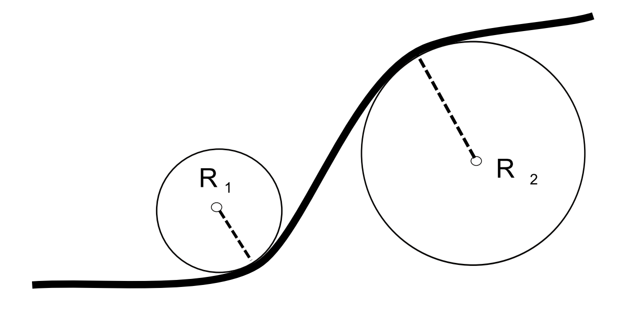


Figura 16.7: La curvatura es el inverso del radio ($1/R$) del círculo que ajusta a la curva en un punto dado. Por convención, es positiva en el caso convexo ($R_2 > 0$) y negativa en el concavo ($R_1 < 0$).

las que el agua experimentará una aceleración. Por el contrario, si esta curvatura es cóncava, el agua tenderá a decelerarse. Las zonas con curvatura vertical cóncava son zonas con tendencia a acumular agua.

Para la curvatura horizontal, se tiene [185]:

$$kh = -\frac{q^2r - pqs + p^2t}{(p^2+q^2)\sqrt{1+p^2+q^2}} \qquad (16.5.28)$$

La interpretación del significado de esta curvatura podemos realizarla de modo similar al caso anterior, en esta ocasión considerando que la sección normal es perpendicular a la línea de máxima pendiente. Una curvatura horizontal convexa representa una zona en la que el flujo tiende a dispersarse, mientras que si es cóncava el flujo tiende a concentrarse, ya que las líneas de flujo convergen.

Junto a estas dos curvaturas más habituales y utilizadas, existen otras también estudiadas en la literatura, si bien con menos aplicación directa por el momento. [181] propone un sistema de 12 curvaturas básicas, incluyendo las anteriores:

- Curvatura horizontal
- Curvatura vertical
- Curvatura media
- Curvatura máxima
- Curvatura mínima
- Curvatura de anesfericidad
- Diferencia de curvatura
- Exceso de curvatura vertical
- Exceso de curvatura horizontal
- Curvatura de acumulación
- Curvatura total acumulada
- Curvatura total anular
- Rotor (curvatura de las líneas de flujo)

A cada uno de estos parámetros puede asignársele una interpretación geomorfométrica, aunque el uso práctico no es frecuente, salvo en el caso de las curvaturas horizontal y vertical.

De igual modo, el estudio de los parámetros geométricos podría continuarse con las derivadas de tercer grado, aunque su aplicación práctica aún no se encuentra bien definida. Es interesante hacer ver, no obstante, que las derivadas sucesivas son más sensibles a los errores del MDE y propagan esta sensibilidad a los resultados obtenidos en base a ellas [168].

Aunque hemos visto una interpretación de las curvaturas relacionada con los flujos de superficie, estas aportan información útil sobre otros

aspectos, tales como la evolución del suelo, pues esta se encuentra directamente relacionada con el comportamiento del agua. También pueden ser útiles, entre otros, para el estudio de las comunidades animales, ya que pueden entenderse las zonas cóncavas como zonas protegidas y las convexas como zonas expuestas, lo cual tiene una indudable repercusión en el desarrollo de dichas comunidades.

16.5.2. Medidas estadísticas

Las medidas estadísticas se basan en la aplicación tanto de conceptos estadísticos básicos como otros más específicos definidos de manera particular para su aplicación sobre un MDE. Son también en su mayoría elementos derivados de un análisis focal, pudiendo en este caso emplear ventanas de análisis de formas y tamaños diversos, al no estar condicionados por el uso de ningún modelo matemático del relieve.

Los parámetros más inmediatos a calcular son los cuatro primeros momentos de la distribución de valores en la ventana de análisis, es decir, la media, la desviación típica, el coeficiente de asimetría y el de kurtosis.

La media es un parámetro genérico que se puede utilizar como un operador focal sobre cualquier capa de datos, y ya la hemos visto en otros apartados anteriores. Tiene, como ya se mencionó (15.2.2), un efecto de suavizar la transición entre los valores de las celdas, produciendo un desenfoque en el caso de aplicarlo sobre imágenes. En el caso del MDE, este suavizado elimina las depresiones compuestas por una única celda.

En algunos casos existe una marcada correlación entre un parámetro estadístico y uno geométrico de los vistos en el punto anterior. Por ejemplo, la desviación típica esta fuertemente correlacionada con la pendiente.

Otra medida de interés es el rango de valores, esto es, la diferencia entre la elevación máxima y la mínima dentro de la ventana de análisis.

La combinación de parámetros estadísticos con parámetros geométricos puede aportar igualmente información de interés. El cálculo de la media aplicada no a la elevación sino a parámetros tales como pendiente o curvatura son ejemplos de lo anterior.

Un parámetro más especifico es la *rugosidad*. Como indica [160], no existe un forma unánimemente aceptada de calcular la rugosidad, ya que los distintos autores aportan soluciones muy diversas. El concepto de rugosidad, no obstante, es sencillo: expresa lo escarpado del relieve, caracterizando la complejidad de este. Terrenos más ondulados o escarpados llevan asociados valores mayores de rugosidad.

La forma más simple de calcular la rugosidad es empleando la desviación típica de las celdas dentro de la ventana de análisis, un parámetro ya comentado. Valores altos de este denotan una superficie irregular, mientras que valores bajos indican un relieve suave. Este método, no obstante, no se comporta correctamente en todas las situaciones, ya que, por ejemplo, asigna un valor alto a una ventana de análisis que engloba una superficie suave —lisa— pero en pendiente. Una solución para solventar esto es ajustar un plano a los valores de la ventana de análisis y calcular la desviación típica de los valores ajustados.

Un enfoque distinto es el propuesto por [175], incluyendo elementos de cálculo vectorial con su *factor de rugosidad*:

$$R = \frac{\sqrt{\left(\sum_{i=1}^{n} \mathrm{X}_i\right)^2 + \left(\sum_{i=1}^{n} \mathrm{Y}_i\right)^2 + \left(\sum_{i=1}^{n} \mathrm{Z}_i\right)^2}}{n} \tag{16.5.29}$$

siendo n el número de celdas en la ventana de análisis y X_i, Y_i y Z_i las componentes del vector

unitario normal a la superficie en cada una de dichas celdas. Estas pueden calcularse a partir de la pendiente y la orientación según las expresiones:

$$\begin{aligned} \mathrm{X}_i &= \sin(s)\cdot\cos(a) \\ \mathrm{Y}_i &= \sin(s)\cdot\sin(a) \\ \mathrm{Z}_i &= \cos(s) \end{aligned} \qquad (16.5.30)$$

Acerca de la rugosidad, y puesto que representa la complejidad del relieve, debe mencionarse la influencia que tiene en la creación del MDE. Relieves rugosos implican una mayor complejidad y son más difíciles de registrar de forma precisa, con lo que la exactitud de las técnicas de interpolación al generar un MDE depende directamente de la rugosidad del relieve a interpolar.

16.6. Análisis hidrológico

Uno de los fenómenos principales que tiene lugar sobre la superficie terrestre es el movimiento de flujos, tanto líquidos como sólidos. El MDE es una herramienta de primer orden para un análisis exhaustivo de estos fenómenos hidrológicos.

Veíamos en el apartado 16.3 cómo resultaba necesario corregir el MDE para eliminar las depresiones de origen artificial y así poder asignar direcciones de flujo en todas las celdas. Trabajando sobre la base de ese MDE corregido estudiaremos a continuación las diversas formas de asignar dichas direcciones y, posteriormente, cómo emplearlas para el cálculo de nuevos parámetros.

La asignación de direcciones de flujo se lleva a cabo mediante un análisis de tipo focal, al igual que los parámetros morfométricos que hemos visto hasta el momento. Esas direcciones, no obstante, implican una conectividad entre las celdas del MDE, y esta conectividad la emplearemos para definir regiones de celdas hidrológicamente relacionadas entre sí y realizar un análisis regional.

16.6.1. Direcciones de flujo

Las direcciones de flujo definen el comportamiento hidrológico de cada celda como unidad aislada. Este comportamiento viene condicionado directamente por las celdas circundantes, ya que es la relación entre las alturas de cada una de ellas con la celda central la que define la dirección de flujo. Dicha dirección está fundamentada en el principio básico y evidente de que el agua en su recorrido se mueve siempre hacia el menor potencial posible, esto es, hacia la dirección de máxima pendiente.

El modelo más sencillo de asignación de direcciones de flujo es el denominado *D8* [186], acrónimo derivado de su denominación en lengua inglesa *Deterministic 8*. Esta formulación es la implementada en la gran mayoría de SIG con capacidades de análisis hidrológico, debido principalmente a su sencillez operativa y facilidad de comprensión.

Los fundamentos teóricos de esta metodología son intuitivos e inmediatos. Dada una celda problema, el flujo puede abandonar esta únicamente en ocho direcciones posibles, esto es, las ocho que resultan de unir el centro de la celda inicial con los de las las ocho circundantes en la ventana de análisis 3×3. El ángulo de desplazamiento del flujo queda así discretizado en intervalos de $45°(360°/8)$.

El proceso de cálculo de la dirección de flujo es sencillo, y requiere únicamente el cálculo de una pendiente entre la celda central y las 8 circundantes, la cual se evalúa mediante la expresión

$$\gamma = \frac{z_5 - z_i}{l}; \qquad z = 1, \ldots, 9; \qquad i \neq 5 \qquad (16.6.1)$$

donde l representa la distancia entre los centros de las celdas entre las que se calcula la pendiente. No se emplea aquí directamente la variable Δs de tamaño de celda, ya que la distancia a aplicar

no es igual en las celdas situadas en la vertical o en la horizontal de la celda central ($l = \Delta s$), que en las celdas situadas en diagonal a la misma ($l = \sqrt{2}\Delta s$). La dirección de flujo viene definida por la celda que marca la mayor de las pendientes.

La principal ventaja del método, su simplicidad, es también su principal defecto. La discretización de las direcciones en múltiplos de 45°hace que flujos que presenten rutas ciertamente distintas (con diferencias en orientación tan importantes como esos mismos 45°) puedan quedar englobados dentro de un misma dirección, ya que la dirección real resulta truncada para encajar dentro de uno de los intervalos.

La aparición de segmentos rectilíneos paralelos en zonas llanas (Figura 16.9) o la acumulación de errores en pendientes monótonas (16.8), son, entre otras, algunas de las consecuencias indeseadas de lo anterior.

Asimismo, existe una imposibilidad para modelizar los procesos de divergencia del flujo, al contemplarse tan solo el caso en que este vierte sobre una única de sus celdas circundantes. Es decir, que el flujo queda modelizado de forma unidimensional, no bidimensional.

Partiendo de la base conceptual del D8, se han elaborado modelos que tratan de superar las deficiencias anteriores y dar lugar a una modelización más realista.

[187] introduce un elemento estocástico en su modelo *Rho8*, manteniendo el resto de conceptos del D8. De este modo, aunque no se elimina la discretización ni la concepción unidimensional del flujo, se soluciona el problema de las líneas de flujo paralelas o la acumulación del error en pendientes uniformes.

En lugar de calcular la dirección de flujo según la máxima pendiente con las celdas circundantes, se calcula la orientación de la celda y en función de esta se asigna hacia una de las dos celdas que definen una dirección más similar. Por ejemplo, para una celda con orientación 15°, la dirección asignada puede ser hacia la celda superior (0°) o la superior derecha (45°). La asignación de una u otra dirección se produce con una probabilidad p establecida en función de la diferencia entre la orientación de la celda y la de cada dirección posible. Así, para el ejemplo mencionado, se tendría

$$\begin{aligned} d_1 = 15 - 0 = 15 \quad &, \quad d_2 = 45 - 15 = 30 \\ p_1 = 1 - \frac{15}{45} = \frac{2}{3} \quad &, \quad p_2 = 1 - \frac{30}{45} = \frac{1}{3} \end{aligned} \tag{16.6.2}$$

Aunque mejora al D8, este método presenta el inconveniente de su aleatoriedad, y se encuentra raramente implementado.

Un modelo más elaborado que tiene en cuenta la divergencia de flujo es el FD8 [188], en el cual el flujo en lugar de verter hacia una única celda, se reparte entre las circundantes. La proporción de flujo asignada a cada una de las 8 celdas vecinas viene dada por

$$f_m = \frac{(\tan s_m)^x}{\sum_{i=1}^{8}(\tan s_i)^x} \tag{16.6.3}$$

donde f_m es el flujo asignado a la celda m de la ventana de análisis, s_i la pendiente hacia la celda i, y x es un factor de convergencia cuyo valor define el comportamiento del algoritmo. A valores más altos de este coeficiente, el flujo resultante es más concentrado, teniéndose una mayor dispersión del mismo para valores bajos. La formulación original emplea un valor $x = 1$, y [189] demuestra que este es un valor óptimo, especialmente en terrenos convexos. Autores como [190] recomiendan el uso de valores entre 4 y 6, considerando que de este modo se obtiene un adecuado equilibrio entre los enfoques anteriormente comentados.

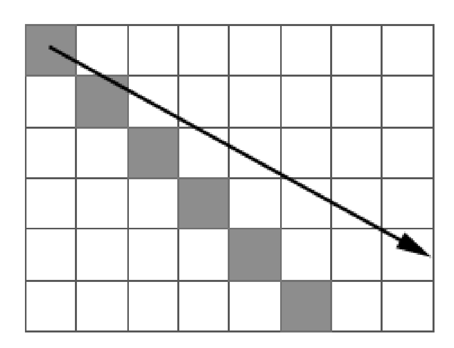


Figura 16.8: Propagación de errores de dirección en el modelo D8.La flecha señala la dirección real de la pendiente y las celdas sombreadas la dirección de flujo calculada

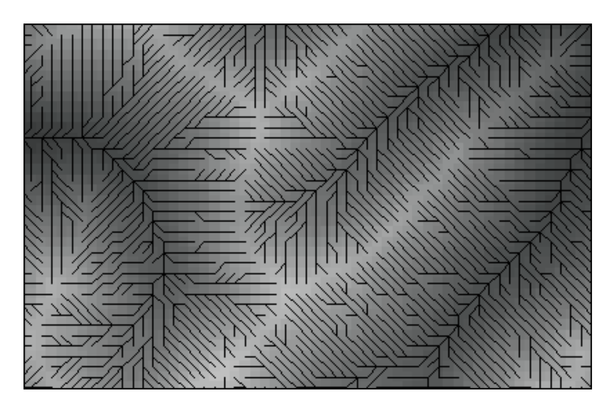


Figura 16.9: Líneas paralelas de flujo como resultado de la aplicación del modelo D8

Por otra parte, los valores de las tangentes pueden obtenerse sencillamente según lo expresado en 16.6.1. Sólo deben emplearse aquellas tangentes que sean positivas, esto es, que representen ángulos entre la celda central y celdas de inferior elevación, pues es exclusivamente hacia estas hacia donde va a desplazarse el flujo.

[191] propone otra solución similar que pretende por un lado modelizar de forma más realista la difusión del flujo, y por otro aportar una formulación más robusta. El modelo propuesto sigue basándose en el D8, solventando el problema de la discretización de flujo en 8 direcciones únicas, y para indicarlo es bautizado como D∞.

Se considera la submalla 3×3 como una superficie representada por caras triangulares, cada uno de los cuales está definido por la celda central y dos celdas consecutivas de su entorno. Se tienen así un total de 8 de estos triángulos, cada uno de los cuales define un plano que permite el cálculo de una pendiente asociada.

Con las pendientes de todos y cada uno de los ocho triángulos, simplemente se tomará la mayor de todas ellas y la orientación asociada a dicho triángulo, siendo esta la que deberá ser considerada de cara a establecer la dirección de flujo en la celda analizada.

Resta, por último, establecer la difusión entre las dos celdas sobre las que se sitúa el triángulo, para lo cual se aplica una sencilla proporción como la siguiente.

$$
\begin{aligned}
f_a &= \frac{\phi_b}{\phi_a + \phi_b} \\
f_b &= \frac{\phi_a}{\phi_a + \phi_b}
\end{aligned}
\qquad (16.6.4)
$$

siendo f_a y f_b las proporciones de flujo que deben asignarse a cada celda de las anteriores, y ϕ_a y ϕ_b las diferencias angulares entre la orientación y las direcciones marcadas por dichas celdas.

Aunque muy poco habituales, existen asimismo metodologías no basadas en el modelo D8, en las cuales el flujo no se desplaza entre los centros de una celda y las de su entorno, sino libremente a través de toda la celda sin restringirse a su punto central (Figura 16.10). El modelo KRA (*Kinematic Routing Algorithm*)[192], de flujo unidimensional, o el DEMON (*Digital Elevation MOdel Networks*) [193], de flujo bidimensional, son ejemplo de ellos.

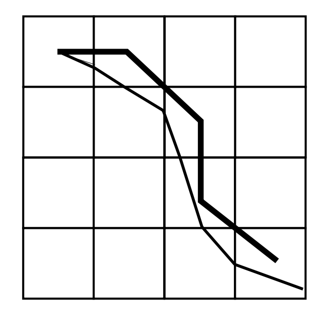


Figura 16.10: Cálculo de dirección de flujo según Kinematic Routing Algorithm (KRA)[192](trazo fino) y comparación con D8 [186] (trazo grueso)

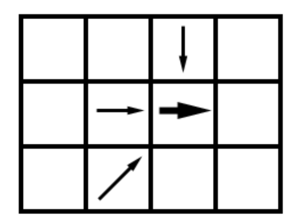


Figura 16.11: Cálculo de dirección de flujo sobre una celda llana (en trazo grueso) como suma vectorial de las direcciones de las celdas que vierten sobre la misma (en trazo fino).

Otro autores como [194] proponen modelos que aplican distintas metodologías en función de la forma del relieve en cada celda, distinguiendo entre grupos de formas orográficas.

16.6.2. Zonas llanas y depresiones

Como ya se dijo antes, las zonas llanas y las depresiones representan un problema para la asignación de direcciones de flujo. Ahora que ya conocemos los algoritmos para llevar a cabo esto último, podemos analizar en profundidad el problema que estos elementos representan y la forma de solucionarlo. Por simplicidad, suponemos que utilizamos el algoritmo D8.

Zonas llanas

Sea la siguiente ventana de análisis:

60	60	60
60	60	60
60	60	60

Puesto que no existe pendiente entre la celda central y las circundantes, no es posible asignar una dirección de flujo a la celda central. Incluso en el caso de no utilizar el D8, tampoco sería posible, ya que otros algoritmos utilizan principios similares o se basan en la orientación, la cual no se encuentra definida para celdas sin pendiente, como vimos en 16.5.14. La intuición y la mera observación nos dicen, no obstante, que el agua se desplazará a través del relieve definido por la anterior ventana.

En realidad, puede considerarse que las zonas planas son un artificio dentro de un MDE, ya que una zona perfectamente plana no existe en la naturaleza como tal. En general, una resolución vertical escasa, bien sea en el almacenamiento o bien en la medición de los valores, es la causante de que estas aparezcan en el MDE.

Supongamos que la anterior ventana se encuentra dentro de un área más amplia con los siguiente valores:

50	60	60	60	70
50	60	60	60	70
50	60	60	60	70

En este caso, resulta fácil intuir que el flujo se desplazará de Este a Oeste, de las zonas de mayor elevación a las de menor. El relieve circundante contiene, pues, información que puede utilizarse para asignar los valores dentro de la zona llana.

Existen dos formas de aplicar estas ideas para dar solución al problema:

- Asignar la dirección de flujo en la zona llana en función de las direcciones en las zonas circundantes donde los algoritmos correspondientes puedan ser aplicados.
- Modificar las elevaciones de la zona llana, añadiéndoles una pendiente que refleje la tendencia del relieve y la transición entre las zonas circundantes de mayor y menor elevación.

Respecto a la primera, podemos asimilar la dirección de flujo de la celda a la tendencia en ese sentido de las celdas que vierten sobre ella. Trabajando sobre la metodología del D8 para ello, y entendiendo cada dirección de flujo como un vector que une el centro de la celda origen con el de aquella sobre la que vierte su flujo, una simple suma vectorial de los vectores que representan dichos flujos dará lugar a un nuevo vector, el cual puede utilizarse como dirección de flujo (Figura 16.11).

Si alguna de las celdas situadas alrededor de la celda problema es a su vez plana, al evaluar la dirección de flujo de esta será necesario recurrir también al algoritmo anterior, continuándose así sucesivamente hasta alcanzar una zona no llana. En ese momento, quedarán definidas todas las direcciones anteriores hasta llegar a dicho emplazamiento desde la celda problema original.

Aunque solucionando eficientemente la dificultad de las zonas planas, esta metodología adolece de algunos problemas de precisión, permitiendo la evaluación de una dirección de flujo en zonas planas pero arrojando en determinadas ocasiones resultados con ciertas deficiencias, especialmente para zonas planas de gran extensión. La posibilidad de existencia de más de un punto de salida o la indeterminación en ciertos casos de la dirección a asignar por existir varias opciones válidas implica la introducción de elementos arbitrarios que conllevan una serie de aspectos negativos. El problema de las líneas paralelas de flujo es en este caso uno de los principales defectos que pueden asociarse con esta metodología.

[195] propone una filosofía opuesta empleando las celdas hacia las que vierte la zona llana y que tengan por sí mismas un flujo bien definido. Con ellas, aplicando un proceso iterativo similar al anterior, se define el flujo de las celdas circundantes a las mismas encaminándolo hacia ellas y continuando de este modo hacia las restantes celdas hasta completar la totalidad de la zona llana (Figura 16.12). No obstante, la similitud en cuanto a sus conceptos, aunque con algunas diferencias, hace que puedan asociarse a esta alternativa unas desventajas similares a las comentadas en el método anterior.

El segundo planteamiento, la modificación del MDE para darle una cierta pendiente sobre las zonas llanas, es más adecuado, puesto que da como resultado un MDE hidrológicamente correcto sobre el que pueden aplicarse sin problemas las formulaciones habituales. [196] propone establecer dicha modificación aplicando los conceptos de las dos soluciones anteriores. Es decir, considerando que el flujo sigue la dirección marcada por las celdas aguas arriba, alejándose de las zonas altas, y confluyendo hacia la dirección señalada por las celdas aguas abajo.

Para ello, define dos capas de elevaciones modificadas, una reflejando cada uno de los fenómenos anteriores (alejamiento de las zonas altas y direccionamiento hacia las zonas bajas), la suma de las cuales da lugar al MDE modificado sin zonas llanas (Figura 16.13).

Depresiones

Sea la siguiente ventana de análisis:

60	60	60
60	58	60
60	60	60

Nos encontramos con un problema similar al anterior, ya que todas las pendientes desde la celda central son positivas, y no existe ninguna dirección de flujo que represente un movimiento hacia aguas abajo. En la realidad, no obstante, una depresión como la anterior no hace que el agua detenga su movimiento. El flujo rellenará la depresión y verterá por la celda de menor elevación, continuando su camino.

Una depresión no siempre lleva implícita la imposibilidad de asignar una dirección de flujo a todas sus celdas. Sea el siguiente fragmento de un MDE.

50	59	60	60	70
50	60	58	57	70
50	60	60	60	70

En este caso, de las dos celdas de la depresión, solo una de ellas no tiene dirección de flujo (de elevación 57), pero en la otra (de elevación 58) esta es incorrecta, puesto que vierte sobre la primera y la tendencia natural del terreno es la contraria. El proceso que tiene lugar será bien distinto, ya que el flujo rellenará la depresión y saldrá de esta desde la celda de elevación 58 y hacia la de menor elevación de su entorno, en este caso 59.

Las depresiones, especialmente las que están formadas por un número pequeño de celdas o incluso una sola celda, son en general producto de los procesos de interpolación que tienen lugar para la creación del MDE. No obstante, pueden deberse a otras razones. Distinguimos según esto los siguientes tipos de depresiones [72].

- Depresiones artificiales debidas a los procesos de creación del MDE
- Depresiones naturales que deben ser eliminadas, tales como lagos o embalses, ya que en realidad estas depresiones se encuentran llenas y se produce un flujo sobre ellas.
- Depresiones naturales que no deben ser eliminadas, tales como las las existentes en zonas kársticas, que suponen efectivamente un impedimento para el flujo.

El procesado del MDE debe trabajar con los dos primeros casos, ya que, independientemente de su origen, es necesario eliminar las depresiones que, al aplicar sobre ellas los algoritmos de flujo, no dan resultados acordes con el proceso real que tiene lugar.

Una primera forma de eliminar las depresiones es mediante el empleo de filtros [197]. Esta metodología, no obstante, no soluciona el problema cuando las depresiones son de varias celdas, y además modifica la totalidad del MDE, incluyendo las celdas correctas que no forman parte de las depresiones.

Las metodologías más elaboradas tratan de simular el verdadero comportamiento del agua, que llena la depresión hasta una altura fija igual en todas las celdas, y vierte así fuera de esta. Autores como [197] suponen que la presencia de una depresión es debida a un error en los valores de elevación, siendo este error siempre por defecto. Es necesario en este caso, por tanto, elevar dichos valores.

Otros como [198] adoptan el enfoque anterior, pero añaden al mismo una posibilidad totalmente opuesta, esto es, que la depresión sea debida a una estimación excesiva de la altura de una celda. Esto provoca que dicha celda (o celdas) actúen a modo de dique, obstruyendo el desplazamiento del flujo a lo largo de las celdas erróneas. La solución en este caso contempla de igual modo la disminución de la cota en aquellas celdas donde se presenten estas características, «abriendo» caminos para el flujo.

Al aplicar alguno de los algoritmos anteriores, obtenemos sobre la depresión una zona plana de cota fija, suficiente para que el flujo pueda conducirse fuera de esta con los algoritmos habituales.

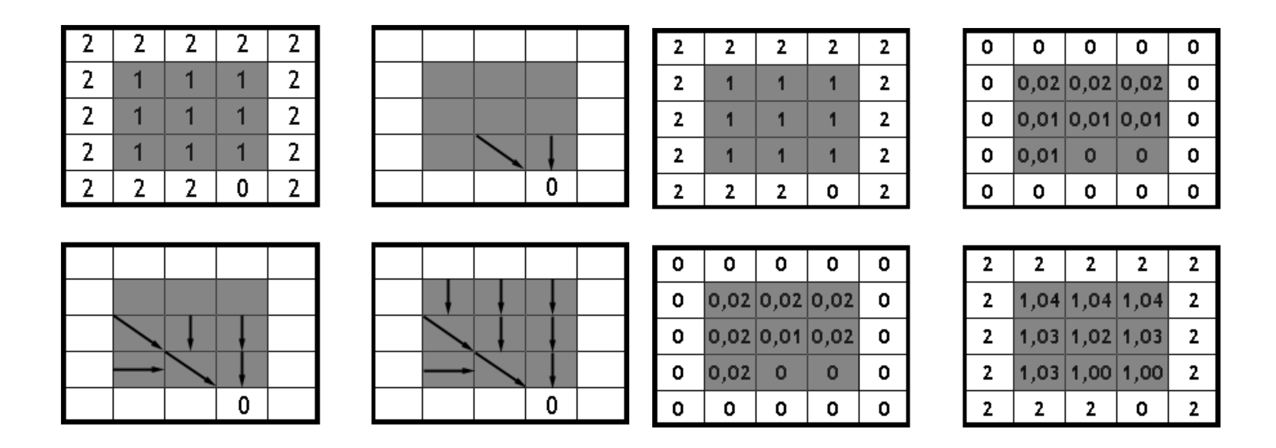


Figura 16.12: Cálculo de direcciones de flujo sobre zonas llanas según [195]. La malla de la izquierda representa las alturas del MDE, donde se aprecia la zona llana de elevación 1 (en gris). Las mallas sucesivas presentan las etapas del proceso iterativo de asignación de direcciones de flujo a partir del punto de salida.

Figura 16.13: Modificación del MDE para el cálculo de direcciones de flujo sobre zonas llanas según [196]. De izquierda a derecha y de arriba a abajo: elevaciones originales, modificación de elevaciones según los dos supuestos considerados y elevaciones finales resultantes.

El MDE resultante puede modificarse nuevamente aplicando los algoritmos ya explicados para la eliminación de zonas llanas, obteniéndose así un nuevo MDE ya hidrológicamente correcto.

La figura 16.14 presenta un esquema comparativo de los diferentes resultados que se obtienen mediante la aplicación de un simple llenado y de la consideración conjunta de llenado y apertura.

[199] propone una solución con un planteamiento radicalmente opuesto. En este caso el algoritmo trabaja «inundando» la totalidad el MDE —elevando la cota de todo él como si hubiera un exceso de agua a lo largo de toda su extensión— y posteriormente retirando dicha capa de agua a excepción de en las depresiones, donde esta se queda y las llena, eliminándolas así.

La figura (16.15) muestra un sencillo esquema del funcionamiento de esta formulación.

16.6.3. Área acumulada y parámetros derivados

El área acumulada es el parámetro principal que puede obtenerse de la aplicación de los algoritmos de direcciones de flujo. Las direcciones de flujo como parámetro focal establecen relaciones entre las celdas y sus vecinas inmediatas. El área acumulada utiliza estas relaciones, pero no en el entorno reducido de la ventana de análisis, sino con carácter zonal, estudiando el conjunto de celdas hidrológicamente conectadas con una dada. En particular, conectadas y situadas aguas arriba de ella.

Para una celda cualquiera, su flujo va a desplazarse por efecto de la gravedad a una o varias (en función del método empleado) de sus celdas vecinas. Estas celdas que reciben el flujo lo desplazaran a su vez a las inferiores, y así sucesivamente. El área acumulada de una celda particular representa el área total de aquellas celdas cuyos flujos, siguiendo este mecanismo, acabaran pa-

2	2	3	0	3	2	2
3	3	3	1	3	2	2
3	2	3	2	3	3	3
3	2	0	0	0	2	3
3	2	0	0	0	2	3
3	2	0	0	0	2	3
3	2	0	0	0	2	3
3	3	3	3	2	2	3
2	2	2	3	3	3	3

2	2	3	0	3	2	2
3	3	3	1	3	2	2
3	2	3	2	3	3	3
3	2	2	2	2	2	3
3	2	2	2	2	2	3
3	2	2	2	2	2	3
3	2	2	2	2	2	3
3	3	3	3	2	2	3
2	2	2	3	3	3	3

2	2	3	0	3	2	2
3	3	3	1	3	2	2
3	2	3	1	3	3	3
3	2	1	1	1	2	3
3	2	1	1	1	2	3
3	2	1	1	1	2	3
3	2	1	1	1	2	3
3	3	3	3	2	2	3
2	2	2	3	3	3	3

Figura 16.14: De derecha a izquierda, elevaciones originales, eliminación de la depresión mediante llenado, y eliminación de la depresión mediante llenado y modificación de celdas obstáculo. En negro, celda de desagüe. En gris, celdas modificadas.

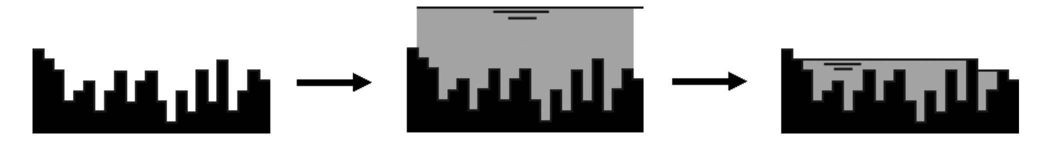


Figura 16.15: Eliminación de depresiones según [199].

sando por dicha celda problema. En el caso del D8, se considera que todo el área de la celda es aportada a la celda inferior. En los métodos de flujo bidimensional, como vimos, solo una fracción es aportada a cada una de las inferiores.

El D8 es, por su simplicidad, el más didáctico de los métodos, y permite comprender visualmente el concepto de este parámetro (Figura 16.16).

La diferencia entre los métodos de asignación de direcciones de flujo se ve de forma gráfica al calcular el área acumulada empleando cada uno de ellos. La figura 16.18 muestra los valores de área acumulada para una superficie generada artificialmente con forma de cono, calculados estos con cada uno de los métodos explicados en 16.6.1.

En la figura 16.17 puede verse una capa de área acumulada calculada a partir de un MDE. Como se dijo en 15.2.1, el uso de una transformación logarítmica enriquece la representación resultante.

Puede aplicarse una ponderación a las celdas, de tal forma que los valores resultantes no expresen un área, sino otra variable distinta. Por ejemplo, puede considerarse un valor de escorrentía para cada celda, de tal modo que se obtendría una escorrentía acumulada.

De igual modo, pueden emplearse algunos de los parámetros que hemos desarrollado anteriormente y calcular después valores medios. Un posible ejemplo es la pendiente media aguas arriba. Esta se calcula con una capa de pendiente acumulada —ponderando las celdas según su pendiente—, y dividiéndola por una capa de área acumulada, expresada esta en numero de celdas. Puesto que refleja la pendiente media de las celdas que aportan flujo, este parámetro puede relacionarse con la velocidad y energía de dichos flujos.

Empleando un enfoque similar basado en el estudio de direcciones de flujo y conectividad hidrológica entre celdas, se obtienen otros pa-

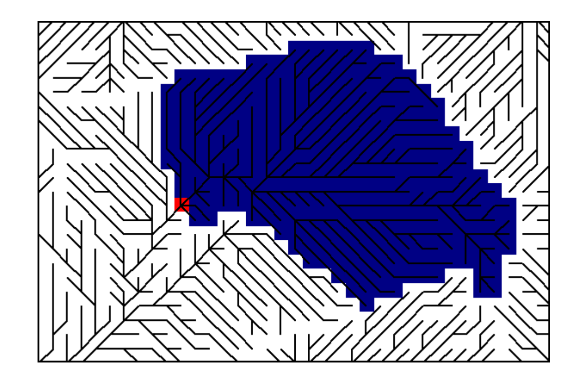


Figura 16.16: El área acumulada de una celda dada (en rojo) es la suma de las áreas de las celdas conectadas con esta y situadas aguas arriba (en azul). Los trazos indican la conectividad entre celdas según el modelo D8

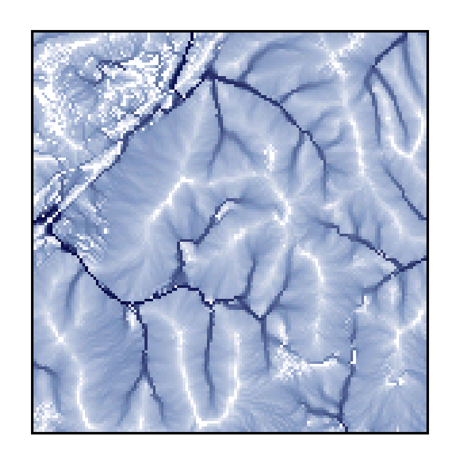
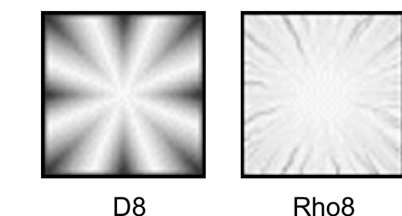
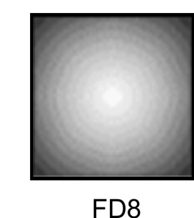
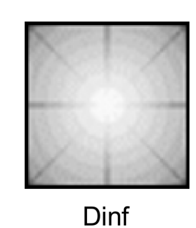
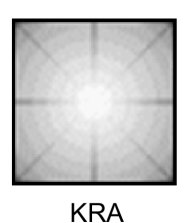


Figura 16.17: Capa de área acumulada. Se ha utilizado una representación logarítmica para aumentar la información visual proporcionada por la imagen.









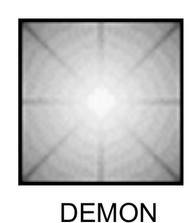


Figura 16.18: Representación del área acumulada sobre un cono circular según los principales algoritmos de asignación de direcciones de flujo (adaptado de [200])

rámetros tales como la longitud de flujo desde aguas arriba. Este valor refleja la distancia máxima recorrida desde cada celda hasta el punto hidrológicamente más alejado de entre los situados aguas arriba de ella. El concepto de orden jerárquico de cauces [201, 202] puede aplicarse a la densa red que conforman todas las conexiones entre celdas, para obtener una nueva capa de ordenes.

Un concepto importante a la hora de calcular el área acumulada u otros parámetros derivados como los anteriores es el de *contaminación de borde*. Se dice que una celda está afectada por contaminación de borde cuando entre aquellas celdas que se encuentran situadas aguas arriba de esta se incluyen celdas en el borde el MDE. Puede suceder que estas celdas de borde tengan aguas arriba de ellas otras celdas, pero, puesto que estas no se encuentran en el MDE, no podemos conocerlas. De esta forma, el valor de área acumulada que se obtiene para las celdas afectadas de contaminación de borde es probable que no sea correcto.

Si tenemos un MDE de la parte baja del cauce, podemos calcular sin problemas todos los parámetros morfométricos tales como pendiente, orientación, etc. Sin embargo, los parámetros hidrológicos no serán correctos, ya que estamos ignorando los tramos medio y alto del cauce, sin

duda con influencia en el tramo bajo que analizamos.

Por último, citar el concepto de *área aportante específica*, que se emplea para el cálculo de índices y otros parámetros derivados. Este área específica se obtiene dividiendo el área entre la anchura de la sección considerada ($a' = a/w$), siendo esta el ancho de celda.

16.6.4. Extracción de redes de drenaje

La extracción de redes de drenaje es una de las tareas principales del análisis hidrológico. El objetivo de este proceso es definir, de entre las celdas del MDE, cuáles forman parte de los cauces y cuáles no. En aquellas que forman parte de los cauces, se asume que existe por tanto un flujo encauzado. En las restantes, el flujo es en ladera.

Para llevar a cabo este análisis se utiliza el propio MDE y una capa de información adicional, sobre la cual se establece una condición que permite el trazado de los cauces. El procedimiento para llevar esto a cabo implica los siguientes pasos:

- Selección de celdas de cabecera, en las cuales se da el inicio de los cauces.
- Delineación de los cauces analizando las celdas aguas abajo de las de cabecera.

Las celdas de cabecera son aquellas para cuales se cumple la condición de existencia de cauce, no cumpliéndose para ninguna de las circundantes situadas aguas arriba.

Respecto a la delineación de los cauces hacia aguas abajo, a pesar de la existencia de otros métodos que hemos ya argumentado como más adecuados para las tareas hasta ahora vistas, se utiliza por regla general el método D8. Con él, se señalan las celdas situadas aguas abajo de las de cabecera, y por la naturaleza del modelo, que supone un flujo unidimensional, los cauces resultantes tiene un ancho de una celda.

Aunque esta forma de proceder es sencilla, presenta una gran numero de posibles soluciones, surgiendo dos cuestiones iniciales: qué capa de información adicional debe utilizarse y qué condición establecer sobre ella.

Como capa de apoyo debemos emplear cualquiera que pueda aportar información relevante sobre la presencia de cauces. El proceso de formación de un cauce depende de numerosos factores tales como la precipitación, o las características del suelo y la litología, pero en última instancia es el relieve quien condiciona la definición de la red de drenaje. Por tanto, es lógico que la capa de información adicional sea una de las derivadas del MDE.

La opción más habitual es emplear el área acumulada como parámetro de apoyo. Zonas con altos valores de la misma implicarán a su vez zonas a las que llega un volumen de escorrentía mayor, ya que este se genera en una superficie mayor aguas arriba. Este mayor volumen hace más probable que el flujo defina en esas celdas un cauce.

Si valores elevados son los que marcan la existencia de celdas de cauce, la condición a imponer consistirá en establecer un umbral y ver qué celdas lo superan. Aquellas que lo superen y cumplan los requisitos para ser celdas de cabecera, serán utilizadas para delinear los cauce hacia aguas abajo de estas.

Aunque sabemos que la condición es de tipo *mayor que*, resta por establecer un valor concreto para el umbral. Del valor escogido depende directamente la red de drenaje resultante. Un umbral mayor hará que existan menos celdas que lo superen, y las cabeceras se situarán en puntos más bajos. La red resultante será menos densa. Por el contrario, si el umbral es menor, las cabeceras se situarán en celdas más elevadas y la red sera más densa y ramificada.

La elección de un umbral debe realizarse de tal modo que la red de drenaje coincida en la mayor medida posible con la realidad fisiográfica, tanto en la coincidencia de las cabeceras con los puntos reales de nacimiento de los cauces como en el numero de estos.

La figura 16.19 muestra tres capas con redes de drenaje extraídas aplicando distintos valores de umbral.

La estimación de un valor óptimo de umbral de forma automatizada es difícil, aunque existen formulaciones al respecto. Quizás la más elaborada de ellas es la desarrollada por [203], quien, fundamentándose en la constancia de las diferencias altitudinales entre los extremos de tramos de un mismo orden [204], propone una aproximación estadística a este cálculo.

Aunque es con mucha diferencia la opción más habitual, la utilización del área acumulada como parámetro de apoyo no es en absoluto ideal. Su utilización implica la generación de redes con densidad de drenaje constante, algo que realmente no sucede en la naturaleza, siendo esta mayor en los tramos altos que en los medios y bajos de los cauces.

[205] propone sustituir el área acumulada por un parámetro de la forma

$$A' = a's^{\alpha} \tag{16.6.5}$$

donde a' es el área aportante específica, s la pendiente en la celda y α un factor que debe determinarse empíricamente, proponiendo el autor para las zonas por él estudiadas un valor de $\alpha = 2$.

[206] propone el uso de la malla de valores de órdenes jerárquicos de Strahler, de tal modo que el establecimiento de un umbral se asemeja a «podar» la estructura arbórea de la red que conforman todas las conexiones entre celdas hidrológicamente relacionadas, hasta reducirla a los tramos de orden superior que forman la red de drenaje real.

Por último, [207] propone utilizar el área acumulada, pero de tal forma que esta refleje solo las celdas aguas arriba que cumplan una condición dada. En particular, se considera que las celdas que tienen una influencia que debe ser considerada sobre el establecimiento o no de cauces y la presencia de fenómenos de flujo encauzado son aquellas de tipo cóncavo, entendiéndose que estas son las que concentran el flujo de agua.

La localización de estas celdas, en lugar de con la aplicación de los conceptos de curvatura vistos en su momento, se lleva a cabo mediante un algoritmo [208] basado en una matriz 2×2 que se desplaza a lo largo de todas las celdas de la malla. En cada punto, de las 4 celdas que componen dicha matriz se señala sobre una nueva malla la celda de mayor elevación. Una vez que se ha desplazado la submalla 2×2 a lo largo del todo el MDE, las celdas que no han sido señaladas representan las celdas de valle cuya influencia debe ser considerada.

La figura 16.20 muestra gráficamente la aplicación del algoritmo de localización de celdas cóncavas.

Con respecto a la extracción de redes de drenaje, existe una preparación del MDE que puede mejorar esta si se dispone de información adicional. Al contrario que la eliminación de zonas llanas y depresiones, no se trata de una preparación necesaria, sino simplemente de configurar el MDE antes de su análisis para reflejar esa información adicional.

Si se dispone de una red de drenaje (por ejemplo, obtenida por digitalización de cartografía clásica de la red fluvial), podemos modificar el MDE para que la red de drenaje que posteriormente extraigamos de él coincida con esa red que conocemos a priori. Este encauzamiento forzado se efectúa a través de la modificación de las cotas de las celdas implicadas.

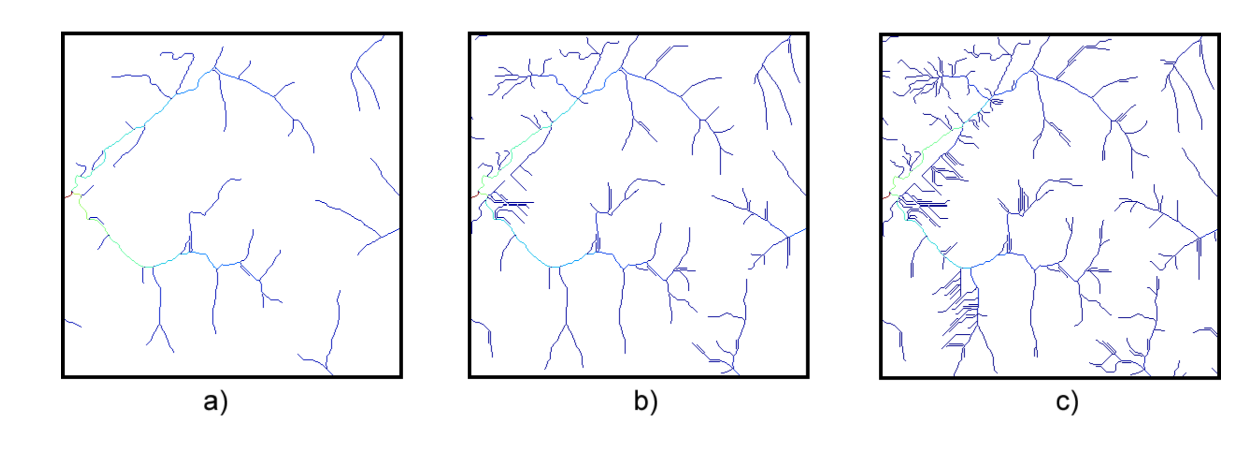


Figura 16.19: Redes de drenaje extraídas para umbrales de area acumulada de 10ha (a), 1ha (b) y 0,1ha (c)

Las celdas a modificar son, en el caso más habitual, aquellas sobre las que se sitúan los cauces que conocemos de antemano. La elevación de estas debe reducirse en un valor fijo h que haga que las celdas adyacentes viertan obligatoriamente sobre ellas.Este proceso se conoce habitualmente como *river burning*[1]

[209] propone una solución más compleja en la que las celdas de cauce se modifican todas ellas en un valor fijo, y aquellas a una distancia de las celdas de cauce menor que un determinado umbral establecido se modifican igualmente, pero este caso en función de dicha distancia. De esta forma, la transición entre las celdas de cauce forzadas y las circundantes se produce de forma suave.

16.6.5. Delimitación y caracterización de cuencas vertientes

Juntos a los cauces, las cuencas vertientes son los otros objetos geográficos con significado hidrológico que pueden extraerse del análisis del MDE. Dada una celda de salida, su cuenca vertiente estará compuesta por todas las celdas aguas arriba de la misma. Estas celdas son las que, sumando su superficie total, nos daban el valor de área acumulada. En este caso, no obstante, el resultado del análisis no es el valor de dicho área, sino el espacio geográfico concreto que ocupa. Además de ésto, para las celdas interiores a la cuenca pueden calcularse valores relacionados con la propia estructura de la cuenca, así como otros que pueden utilizarse para caracterizar esta y realizar un análisis hidrológico más exhaustivo.

Delimitación. Cuencas y subcuencas

Con el conocimiento de las direcciones de flujo y las conexiones entre celdas, el procedimiento para delinear una cuenca vertiente no difiere mucho del cálculo del área acumulada para una celda. Al igual que sucedía en la extracción de cauces, lo general es aplicar el método D8, ya que da lugar a cuencas bien definidas, que pueden representarse mediante capas ráster con dos únicos valores, uno para las celdas que pertenecen a la cuenca vertiente y otro para las que no. Si se aplica otra metodología tal como D∞ o FD8, al dividirse el flujo entre las celdas circundantes, pueden existir

[1] Literalmente, *quemar los rios* sobre el MDE, ya que es como si su forma fuera marcada a fuego sobre este.

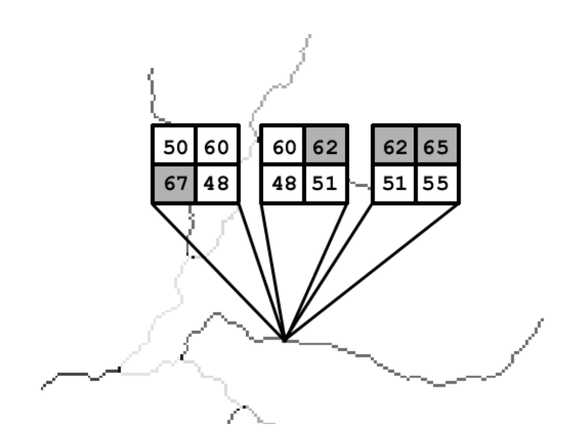


Figura 16.20: Identificación de celdas de valle según [208]. En cada pasada se señala la celda más elevada de cada cuatro. Las celdas no señaladas al final del proceso constituyen las celdas de valle.

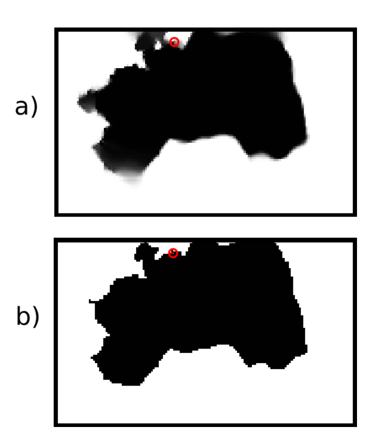


Figura 16.21: Comparación entre una cuenca calculada mediante el método FD8 (a) y otra aplicando el método D8 (b). Las celdas en color gris solo aportan parte de su flujo a la unidad hidrológica. El punto rojo señala el punto de cierre.

celdas que aporten su flujo a la cuenca pero no de forma completa, con lo que su pertenencia a la misma no lo es en el mismo grado que las restantes. La figura 16.21 muestra gráficamente la diferencia entre las cuencas obtenidas utilizando los métodos D8 y FD8 respectivamente.

Nótese que, aunque ambas cuencas sufren el efecto de la contaminación de borde, la calculada por el método FD8 lo hace en mayor medida. Es decir, el algoritmo de flujo puede condicionar la extensión de la cuenca vertiente y con ello efectos tales como el de contaminación de borde.

Por medio de un análisis similar al anterior podemos no solo calcular la cuenca a un punto dado, sino subdividir esta en subunidades hidrológicas menores. El conjunto de estas subunidades conforma una teselación del espacio ocupado por la cuenca en subcuencas independientes.

Existen dos formas principales de subdividir una cuenca, requiriéndose información adicional además de la localización de la celda de salida:

- Mediante celdas de salida adicionales correspondientes a cada subcuenca
- Mediante un umbral de tamaño máximo de las subcuencas.

Para el primer caso, se calculan las cuencas vertientes a todas las celdas de salida especificadas, considerándose siempre que la cuenca asociada a las celdas situadas aguas abajo engloba a la de aquellas situadas aguas arriba de ella, y descontando la superficie de intersección. Los puntos de salida generalmente se establecen sobre las celdas que representan confluencias de cauces donde el orden jerárquico de la red de drenaje varía. Este es un ejemplo directo de aplicación de la red de drenaje que extrajimos en el punto anterior.

Si asignamos ordenes jerárquicos de Shreve [210], tendremos una subcuenca para cada subtramo, ya que estos ordenes varían en cada confluencia. Si, por el contrario, empleamos ordenes de Strahler [202], solo serán consideradas como celdas de salida aquella confluencias en las que se unan cauces de igual orden.

El otro método de subdivisión no requiere el empleo de otras celdas de salida además de

la correspondiente a la cuenca global. Las subcuencas se establecen de tal modo que todas ellas tienen un superficie similar, la cual condiciona el número de estas que aparecen.

Caracterización

La información sobre la superficie y forma de la cuenca, así como la de los valores interiores de la misma y las relaciones hidrológicas entre sus celdas, son la base para elementos de caracterización de la cuenca vertiente.

En primer lugar, tenemos el área y perímetro de la cuenca, parámetros clásicos sencillos de calcular sobre la capa que la define, bien contando las celdas o las celdas de borde. El centro de masas del conjunto de celdas es el centro de masas de la cuenca, de interés para el cálculo de parámetros accesorios con significado hidrológico.

Como ya vimos en 15.2.4, la capa con la cuenca puede usarse como máscara para limitar las celdas con valores válidos de una determinada variable a aquellas dentro de la cuenca. Haciendo esto, el histograma de frecuencia de estos valores o una curva hipsográfica calculada con ellos sirve como descriptor de la cuenca.

Un parámetro de interés es la distancia de flujo desde cada celda hasta el punto de salida. Suponiendo una velocidad media en las celdas de la cuenca, estas distancias pueden convertirse en tiempos de salida. La estimación de la velocidad constante puede hacerse, por ejemplo, utilizando el tiempo de concentración de la cuenca. El histograma de frecuencias de estos puede emplearse para obtener un hidrograma unitario de la cuenca (Figura 16.22).

El cálculo puede hacerse de forma más precisa si se suponen distintas velocidades en los cauces y fuera de ellos. [211] propone una relación de la forma

$$V_{ladera} = \frac{V_{cauce}}{k} \qquad (16.6.6)$$

donde k se sitúa según el autor en el intervalo de valores 10−15.

Se puede plantear un modelo más detallado si se consideran las condiciones particulares de cada celda (pendiente, área acumulada, rugosidad hidráulica, etc.), y se asigna una velocidad individual a cada uno en función de ellos. En [212] puede encontrarse un modelo de estas características en el cual las celdas se dividen en tres grupos en función de su área acumulada, aplicándose formulaciones distintas para cada uno de ellos.

16.6.6. Índices hidrológicos

Con todos los parámetros, tanto hidrológicos como morfométricos, que hemos visto hasta el momento, disponemos de una gran cantidad de información acerca del relieve. Combinando estos se pueden crear nuevos índices hidrológicos que caractericen de forma más precisa el comportamiento hidrológico de cada celda. Tres son los principales índices.

- Índice topográfico de humedad
- Índice de potencia de cauce
- Factor topográfico LS

Todos ellos se basan en la pendiente y en el área acumulada específica como parámetros básicos

El *índice topográfico de humedad* [213] se define como

$$I = \ln \frac{a'}{\tan s} \qquad (16.6.7)$$

donde a' es el área acumulada específica y s la pendiente.

Es habitual sustituir la pendiente nula por una de valor muy pequeño, del orden de 0,0001 %

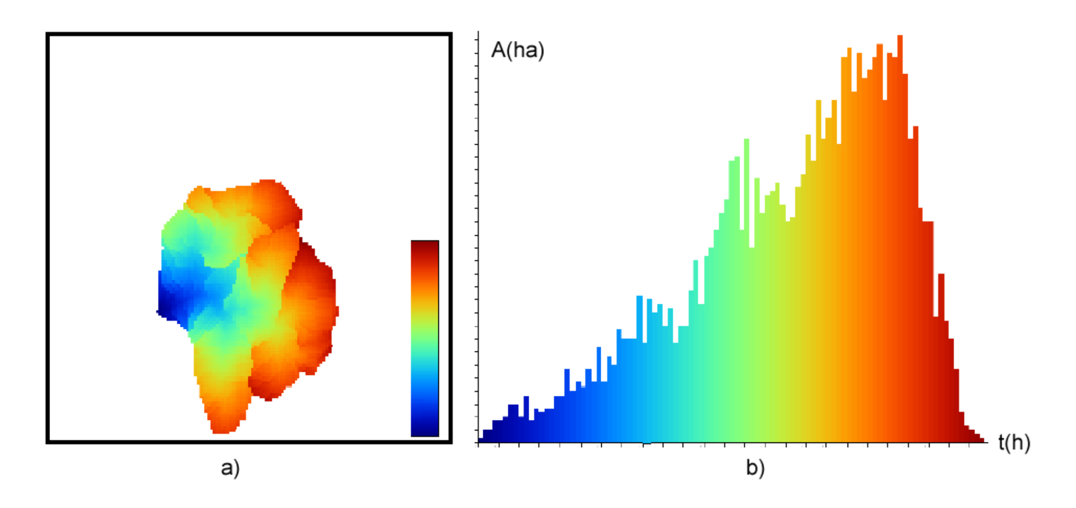


Figura 16.22: a) Capa de tiempos de salida suponiendo velocidad constante. b) Histograma de frecuencias asociado.

por ejemplo, para evitar que el índice no quede definido en celdas planas.

Detrás de su aspecto sencillo, este índice esconde una gran cantidad de información. Originalmente fue definido con objeto de expresar la relación entre la morfología del relieve y la acumulación de agua en el suelo, responsable esta a su vez de la escorrentía generada, velocidad de flujo, etc. En particular, un uso principal de este índice es el relacionado con la producción de escorrentía bajo el supuesto de que esta se produce en presencia de una saturación de agua en el suelo, momento en el que nivel freático alcanza la superficie.

Valores altos del índice implican zonas con baja pendiente o con gran cantidad de área acumulada. Por tanto, se trata de celdas a las que llegarán flujos abundantes y que no evacuan bien dicho flujo, teniendo tendencia a acumular humedad. Por el contrario, las celdas con valores bajos, con elevada pendiente y baja área acumulada, serán celdas en las que no se encontrará humedad edáfica.

De formulación similar el anterior, el *índice de potencia de cauce* caracteriza la energía de los flujos a través de cada celda. Su expresión es la siguiente:

$$P = a's \tag{16.6.8}$$

Valores altos tanto de área acumulada como de pendiente, ambos implican una mayor energía del flujo. En un caso, porque este flujo sera abundante, y en otro porque llevará una mayor velocidad.

Por último, el factor LS de la Ecuación Universal de Pérdidas de Suelo, originalmente basado en la longitud de la pendiente y su magnitud, se sustituye por la siguiente expresión [214], empleando el área acumulada específica:

$$LS = \left(\frac{a'}{22,13}\right)^m \left(\frac{\sin\beta}{0,0896}\right)^n \tag{16.6.9}$$

donde m y n son parámetros a establecer empíricamente, para los cuales [214] demuestran que, con laderas menores de 100 metros de longitud y pendientes hasta 14°los valores $m = 0,6$ y

$n = 1,4$ dan resultados ajustados a las formulaciones originales. Por su parte, [215] propone unos valores de $m = n = 1$ para el caso en que la erosión predominante sea de tipo laminar, o bien $m = 1,6$ y $n = 1,3$ en caso de que exista predominio de erosión en regueros.

El hecho de emplear área acumulada en lugar de la longitud de pendiente hace que se tengan en cuenta factores como la convergencia de flujo, no contemplados en la formulación original.

16.7. Visibilidad

Las formas del relieve condicionan la visibilidad entre puntos del MDE, pudiendo bloquear la línea de visión entre ellos. Un análisis de visibilidad permite saber si dos celdas tienen conexión visual o conocer qué otras celdas son visibles desde una celda dada. Puesto que el concepto de visibilidad es recíproco, esto es equivalente a calcular qué otras celdas ven a una celda dada.

El conjunto de celdas relacionadas visualmente con una concreta conforman la *cuenca visual* asociada a esta. El cálculo de la cuenca visual implica el estudio de todas las direcciones en las que la luz llega a (o parte de) la celda, analizando para cada una de ellas una *linea de visión*.

Cada línea de visión pasa por una serie de celdas, las cuales serán visibles o no en función de las formas del relieve que se interpongan entre estas y la celda central que define la cuenca visual. Para calcular si cada una de estas celdas A es visible desde la central B, calculamos el ángulo de la línea que las une, según

$$\alpha = arctan\left(\frac{z_b - z_a}{d_{AB}}\right) \qquad (16.7.1)$$

siendo d_{AB} la distancia entre A y B. Si el ángulo formado por cualquier otra celda, B', más próxima a A que B es mayor, entonces B no es visible desde A. El procedimiento puede ilustrarse con un pequeño ejemplo numérico como el siguiente.

Sobre la capa ráster de la figura 16.23, definimos una línea de visión entre las celdas superior e inferior de la columna izquierda. Si suponemos un tamaño de celda igual a 1, los valores de los ángulos y si son visibles o no desde la primera celda (la de la esquina superior izquierda) se presentan en la tabla 16.1.

Aunque existen similitudes entre el concepto de una cuenca hidrológica y una cuenca visual, esta última no ha de estar necesariamente formada por un único polígono, sino que puede componerse de varios polígonos aislados entre sí. Igualmente, no existe una relación entre las celdas de la cuenca (no han de verse entre sí), al contrario de lo que ocurre en la cuenca hidrológica donde sí se da esta relación.

En el cálculo de visibilidades, ha de incorporarse asimismo el significado físico del fenómeno analizado. Dos puntos a una distancia de muchos kilómetros, si no existe relieve entre ellos, resultarán como visibles entre sí si efectuamos un análisis como el descrito. No obstante, por las propias limitaciones de la visión humana, así como por efectos atmosféricos y otros factores, no es lógico pensar que tales puntos puedan verse. El análisis visual debe, por tanto, limitarse a una distancia acorde con el proceso modelizado.

En la figura 16.24 puede verse la cuenca visual asociada a una celda concreta.

El uso de cuencas visuales es habitual, por ejemplo, para el emplazamiento de antenas de telefonía o radio. Puesto que la intensidad de la señal decrece con la distancia, también en este caso deben delimitarse las cuencas visuales (que en este caso señalarían aquellas celdas que reciben y pueden utilizar la señal de la antena) no en base únicamente a los elementos del relieve, sino también a la distancia.

10	16	23	16	9	6
14	11	18	11	18	19
19	15	13	21	23	25
20	20	19	14	38	45
24	20	20	28	18	49
23	24	34	38	45	51

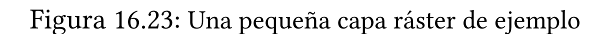


Figura 16.23: Una pequeña capa ráster de ejemplo



Figura 16.24: Cuenca visual asociada a un punto dado (en rojo). Se ha supuesto una altura del objeto de 20 metros.

Fila, col.	H	ΔH	$\Delta H/Dist.$	**Visible sí/no**
1,2	14	4	4	Visible
1,3	19	9	4.5	Visible
1,4	22	12	4	No visible
1,5	24	14	3.5	No visible
1,6	23	13	2.6	No visible

Cuadro 16.1: Análisis de visibilidad sobre una línea de visión.

Partiendo de este razonamiento, podemos extender el concepto de visibilidad, y generar capas con más información. En lugar de generar una nueva capa donde las celdas tengan únicamente dos posibles valores —visible o no visible—, puede asignarse a las celdas visibles valores tales como la distancia al emisor, el ángulo entre emisor y receptor, o el tamaño relativo con que se visualiza el punto. Este último análisis es de utilidad, por ejemplo, para el emplazamiento de elementos constructivos con una altura dada h, tales como aerogeneradores, y evaluar así el impacto visual que producen. Se calcula no solo el número de celdas que ven el elemento —la cantidad de celdas afectadas—, sino *cómo* ven a ese elemento —cómo es esa afección.

Para calcular el tamaño relativo de un objeto de dimensión h visto desde una celda concreta, se utiliza la expresión

$$\mathsf{TAM} = \arctan\left(\frac{h}{d_{AB}}\right) \tag{16.7.2}$$

siendo d_{AB} la distancia entre la celda donde reside el objeto y la celda del observador.

Esta altura que se añade a las celdas donde emplazamos un objeto, en realidad también puede añadirse a las restantes, ya que el observador también tiene una altura dada. En general, no obstante, tiende a no utilizarse esta. Nótese que añadir una altura a un objeto cambia en cierta medida el concepto de visibilidad, ya que, según

el caso analizado, implicara ver el objeto en su totalidad, y no solo la parte superior que marca la altura máxima.

Hasta este punto, hemos utilizado el concepto de visibilidad para un único punto, pero podemos repetir el análisis para todas las celdas o para un conjunto de ellas. Este procedimiento permite calcular otros parámetros, tales como el numero de celdas que se ven desde cada una, pudiendo incluso ponderar estas para adjudicar un significado más concreto a estos valores. Por ejemplo, si ponderamos las celdas en función de su valor paisajístico, el numero de celdas vistas sera un indicador de la belleza escénica de la misma.

El resultado de este análisis de visibilidad repetido se conoce como *exposición visual* [174]. La figura 16.25 muestra el numero de celdas de cauce vistas desde las restantes, sirviendo así para evaluar en cierto modo la belleza paisajística que deriva de la presencia del cauce. Estas celdas a su vez se han ponderado en función de la importancia del cauce.

Relacionado con la exposición visual, encontramos otros análisis derivados, de gran importancia en muchos estudios. Uno de ellos es el análisis de cobertura, que pretende calcular las formas óptimas de cubrir un terreno, de forma que un numero mínimo de observadores pueda observar la totalidad de un área. Dos son los análisis principales en este sentido: determinar el número mínimo de puntos y su localización para que todo un área quede bajo observación, o determinar la localización de un número n de puntos que hace máxima el área observada.

Análisis como estos son útiles para determinar dónde establecer torres de vigilancia de incendios o repetidores de telefonía móvil, entre otros, maximizando la eficiencia de los mismos.

Si los análisis de exposición visual ya implican un gran número de cálculos y requieren mucho tiempo de proceso, los análisis de cobertura añaden elementos combinatorios que hacen que sea imposible tratar el problema mediante un análisis por fuerza bruta (comprobando las posibles soluciones y eligiendo simplemente la mejor). Por ello, estos análisis a menudo utilizan métodos heurísticos o se basan en Modelos Digitales de Elevaciones en formato TIN, ya que en ellos el análisis de visibilidad es menos costoso en términos de número de operaciones necesarias (el número de vértices de un TIN es sensiblemente menor que el número de celdas de una capa ráster). En [216] pueden encontrarse más detalles al respecto.

16.8. Caracterización de formas del terreno

Las distintas formas del terreno pueden clasificarse en base a parámetros morfométricos o mediante análisis locales similares a los empleados para obtener estos.

El número de clases distintas y su significado varía en función del tipo de clasificación y el método, existiendo gran diversidad tanto en la propia clasificación como en la metodología.

Con independencia del método, el tamaño de la ventana de análisis medido en unidades sobre el terreno (no en número de celdas) es de vital importancia para obtener resultados coherentes. Como ya quedo claro en la sección 11.2.1, un mismo relieve puede ser caracterizado de distintas formas en función de la escala a la que se analice. Es por ello que resulta clave la elección de una escala de análisis que se corresponda con el estudio del relieve que pretendemos realizar, ya sea este a escala de microtopografía o de macrorelieve.

Una primera clasificación la podemos obtener a partir del análisis de curvaturas. En particular, tomando las curvaturas horizontal y vertical, podemos combinar estas y definir nueve grupos

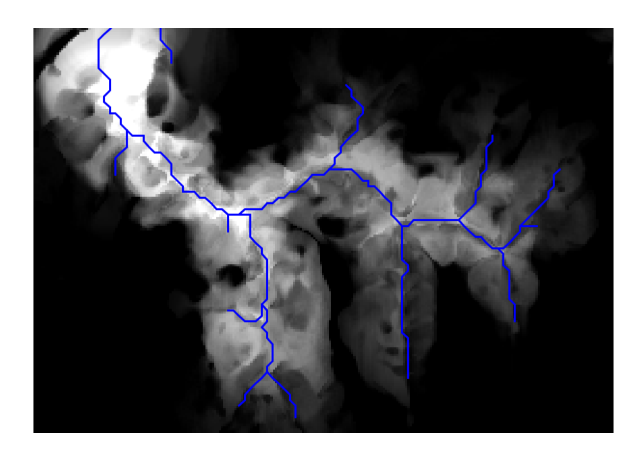


Figura 16.25: Capa de exposición visual de un cauce, reflejando el número de celdas pertenecientes a este que son visibles desde cada una de las restantes, ponderadas por su importancia jerárquica en la red de drenaje.

distintos que caracterizan los procesos de acumulación [217]. Se establece un valor mínimo en valor absoluto, por debajo del cual las celdas se consideran sin curvatura, es decir, planas. Los valores continuos de curvatura pueden ahora categorizarse en tres clases: convexa, cóncava o plana. Cruzando estos tres grupos para las dos curvaturas, obtenemos el conjunto de las nueve clases posibles (Figura 16.26).

Recordando el significado de las curvaturas, tendremos procesos de acumulación en las celdas con ambas curvaturas concavas, ya que el flujo se concentrará y tenderá a frenarse. Del mismo modo, las celdas con ambas curvaturas convexas registrarán procesos de dispersión. Entre estas, las combinaciones restantes identifican zonas de tránsito con diversas características.

Con un planteamiento similar, [151] propone una división en 6 formas del relieve a partir de los valores de las segundas derivadas.

Otra metodología distinta y muy aceptada es la propuesta por [218]. Con anterioridad a la aparición de los Modelos Digitales del Terreno, [219] estableció una clasificación del relieve en base a tres parámetros: la pendiente, el relieve relativo y la forma del perfil. El relieve relativo es el rango de valores de elevación en un entorno dado del punto, y el tipo de perfil se calcula con la distribución de pendientes en los puntos circundantes. Esta clasificación da un total de 96 clases posibles, derivadas de la división de los anteriores parámetros base en cuatro, seis y cuatro clases respectivamente, así como la combinación posterior de los mismos.

[218] adapta este planteamiento al empleo de MDE, de tal modo que se obtienen de forma automática mapas de formas del relieve a partir de estos. Estableciendo una ventana de análisis de un tamaño fijo (9.6 km originalmente), el parámetro relativo a la pendiente se obtiene mediante el recuento de celdas en la ventana que superan un umbral dado (8 % en dicha formulación original). El valor de este recuento, expresado en porcentaje de celdas totales de la ventana, se divide en cuatro clases.

- menos del 20 % de celdas con pendiente mayor que el umbral.
- entre el 20 y el 50 % de celdas con pendiente mayor que el umbral.
- entre el 50 y el 80 % de celdas con pendiente mayor que el umbral.

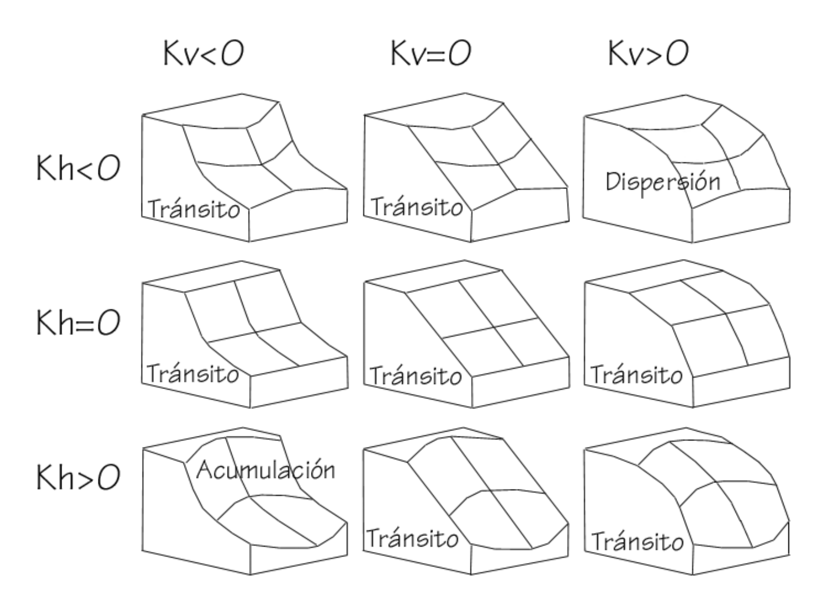


Figura 16.26: Caracterización de procesos de acumulación según los valores de las curvaturas horizontal y vertical asociadas

- más del 80 % de celdas con pendiente mayor que el umbral.

Para el caso del relieve relativo, se establecen asimismo seis grupos, a aplicar sobre el rango de valores en la misma ventana utilizada para el parámetro anterior:

- 0 – 30 m
- 30 – 91 m
- 91 – 152 m
- 152 – 305 m
- 305 – 915 m
- más de 915 m

Por último, se define el tipo de perfil viendo si las celdas con pendiente mayor que el umbral tienen elevación mayor o menor que la central. Se definen las siguientes 4 clases.

- menos del 25 % de celdas con pendiente mayor que el umbral situadas en zonas de mayor elevación.
- entre el 25 % y el 50 % de celdas con pendiente mayor que el umbral situadas en zonas de mayor elevación.
- entre el 50 % y el 75 % de celdas con pendiente mayor que el umbral situadas en zonas de mayor elevación.
- más del 75 % de celdas con pendiente mayor que el umbral situadas en zonas de mayor elevación.

Un enfoque bien distinto es el basado en clasificación no supervisada. La formación del relieve es un proceso que depende de muchos factores, no únicamente los relativos a la morfometría local. Combinando algunos de los parámetros que hemos visto hasta el momento, pueden establecerse clasificaciones a partir de técnicas de clasificación no supervisada (las estudiaremos en el apartado 21.2.2). Por ejemplo, [220] emplea como parámetros la propia elevación, la pendiente, las curvaturas horizontal y vertical, el índice topográfico y la radiación solar. Estableciendo un número de clases a definir, se obtienen una categorización de las celdas de acuerdo con las formas del relieve que representan.

De igual modo, puede plantearse un análisis similar pero utilizando clasificación supervisada [221], o emplear no una clasificación discreta, sino una basada en el uso de lógica difusa [222] [220] [223]. Veremos algunas ideas sobre lógica difusa más adelante en este libro, particularmente en el apartado 21.4.1.

16.9. Resumen

El MDE es una capa de información fundamental, y de él pueden extraerse un número muy elevado de nuevos parámetros. El modelo ráster de malla regular es el más adecuado para ello, aunque debe prepararse para que los resultados derivados sean correctos, en especial los de corte hidrológico.

El análisis del MDE mediante funciones focales se puede llevar a cabo mediante parámetros estadísticos, así como mediante parámetros geométricos. Estos últimos requieren una caracterización matemática del relieve, para poder sobre ella aplicar las herramientas del calculo diferencial. Los parámetros basados en derivadas de primer orden son la pendiente y la orientación, así como, en función de estos, el estudio de la insolación, que puede extenderse mediante conceptos de visibilidad. Los basados en derivadas de segundo grado son las curvaturas, entre las cuales destacan las denominadas curvatura horizontal y vertical.

Dentro del análisis hidrológico, existen diversos métodos para la asignación de direcciones de flujo. El método D8 es el más elemental de ellos, y constituye la base conceptual de gran parte de los restantes. Estos nos permiten establecer relaciones entre las celdas y proceder así a un estudio regional. El área aportante es el parámetro principal de dicho análisis, y se puede emplear como parámetro de apoyo para la extracción de redes de drenaje.

La combinación de parámetros, en particular área aportante y pendiente, da lugar a índices hidrológicos de gran interés.

Procesado de imágenes

17.1. Introducción

Las imágenes como tipo de dato son parte de los SIG desde prácticamente su origen, y las formulaciones dedicadas al análisis de estas existen también desde entonces y son un campo muy desarrollado. Este análisis representa una de las áreas más ricas del SIG, ya que combina las formulaciones que consideran la naturaleza de esta clase particular de imágenes —esto es, el hecho de que son imágenes de la superficie terrestre— junto con las que existen para el análisis y proceso de imágenes digitales en general. Estas últimas constituyen por sí solas un amplio grupo de técnicas que, en conjunto con las dedicadas al análisis de datos espaciales, forman una serie de herramientas de gran utilidad para el uso genérico de los SIG.

Si la consideramos simplemente como una estructura de datos, una imagen no es sino una capa ráster más. Tiene la estructura regular de una capa ráster y contiene en cada celda (cada píxel de la imagen) el valor asociado a una variable dada. Particularmente, esta variable es la intensidad de la reflectancia en una longitud de onda dada.

No existe diferencia «formal» entre un MDE creado a partir de la información recogida por un sensor remoto tal como los utilizados, por ejemplo, para la Shuttle Radar Topographic Mision (SRTM), y una imagen de satélite tal como las que proporciona el satélite LANDSAT (salvo el hecho de que esta última contiene varias bandas en lugar de una única). No obstante, la naturaleza de la variable recogida en la imagen y la componente visual intrínseca a la misma hacen que el campo del análisis de imágenes presente sus particularidades, que serán las que veremos a lo largo de este capítulo.

En conjunto, estudiaremos los elementos básicos del análisis de imágenes digitales, así como las formulaciones específicas para el caso de imágenes aéreas o procedentes de teledetección, es decir, las formulaciones específicas para el tipo de análisis que se realiza en un SIG.

Tanto uno como otro de estos grupos son muy extensos, y en el espacio de este capítulo no pueden recogerse sino las formulaciones más habituales y útiles de ambos. La teledetección queda en muchas ocasiones fuera de los textos al uso sobre SIG, como una disciplina con entidad propia, relegándose también junto a ella al procesado de imágenes. Aun así, los SIG no pueden entenderse hoy día sin el concurso de imágenes, ya sea como elementos visuales o como objetos de análisis, y las metodologías dedicadas a su procesado deben recogerse al menos en cierta medida en un libro como este.

Debido a lo extenso de la materia a tratar y a la fuerte componente matemática con la que puede presentarse, se ha simplificado la exposición, incluyendo estos elementos matemáticos

en limitadas ocasiones y dando más peso al aspecto práctico. En caso de que el lector esté interesado, las referencias indicadas a lo largo del capítulo contienen descripciones detalladas de las fórmulas y procesos que definen cada una de las operaciones sobre imágenes que veremos a continuación

17.1.1. La utilidad de las imágenes en un SIG

Antes de estudiar las técnicas de procesado de imágenes, debemos conocer la utilidad que estas tienen cuando son incorporadas a un SIG.

En líneas generales, podemos identificar los siguientes procesos, que serán a los que se adapten las técnicas que veremos a lo largo del capítulo.

- **Representación**. Las imágenes son, ante todo, elementos visuales, y en muchos casos no se utilizan más que para conformar una base sobre la que emplazar otras capas. En este sentido, las imágenes son utilizadas para proveer un contexto visual en el entorno de trabajo de un SIG.
  Este libro contiene una parte entera dedicada a la generación cartográfica, en la cual se tratan aspectos relativos a la representación de todo tipo de capas de datos espaciales, entre ellas las imágenes. Junto a ello, la literatura es abundante en algoritmos que permiten mejorar las capacidades que una imagen tiene de transmitir su información o de establecer ese citado contexto visual, y estos algoritmos se detallarán en este capítulo.
- **Identificación de elementos**. La realidad que se refleja en una imagen esta compuesta por los objetos presentes sobre la superficie terrestre. Arboles, carreteras, zonas de distinta vegetación, así como otros muchos elementos de pequeño o gran tamaño son los que constituyen esa realidad. Conocemos ya fórmulas para tratar con esos elementos y estudiarlos dentro de un SIG. Obtener a partir de las imágenes nuevas capas que los contengan es una tarea habitual.
  La identificación y delineación de estos elementos puede implicar la creación de nuevas capas vectoriales o bien de capas ráster, así como las transformaciones encaminadas a que dichas operaciones de identificación se realicen de manera lo más precisa posible.
- **Identificación de características**. Si tomamos una imagen, podemos identificar por sus formas una carretera y decir si es ancha o estrecha, larga o corta, o bien si es sinuosa o rectilínea. Estos parámetros geométricos son los que nos ayudan a identificar el elemento en sí, según lo comentado en el punto anterior, pero no es solo lo referente a la geometría lo que una imagen puede contener acerca de esa carretera. Mirando esa misma imagen, es probable que podamos decir por su color si el firme es asfaltado o, por el contrario, es de tierra y se trata de un camino. Esta identificación de las propiedades de un elemento, o bien de una simple celda aislada, es otra de las funciones básicas que se llevan a cabo a partir de imágenes.
  Dentro de este grupo merecen especial mención aquellas formulaciones que estiman propiedades físicas de los elementos recogidos en la imagen. Con los valores de una imagen puede estimarse desde el vigor vegetativo de una masa arbórea hasta la turbidez del agua de un embalse, entre otros muchos parámetros físicos de diversa índole.

17.1.2. Tipos de procesos con imágenes

Teniendo en consideración las anteriores utilidades básicas de las imágenes en un SIG, las

operaciones sobre estas pueden dividirse en tres grupos principales:

- **Corrección**. Los equipos empleados para recoger las imágenes pueden incorporar errores, ruido, o distorsiones. Eliminar o tratar estas para que su efecto sea menor es un proceso previo que en la mayoría de los casos es un requerimiento imprescindible antes del análisis.
- **Mejora**. La mejora es el paso siguiente a la corrección. Una vez que las imágenes se han corregido, la información que contienen puede hacerse más explícita a través de distintos tipos de modificaciones. En el análisis visual, la mejora de una imagen hace que sea más sencillo percibir el contenido de esta y darle un uso directo. Las características o la forma de esa carretera de la que hablábamos pueden verse más claramente. Cuando estos procesos de identificación de elementos y características se realizan no de forma manual, sino automática, también es importante una preparación de las imágenes para mejorarlas de cara a este proceso. Mejorando una imagen, mejoramos también la capacidad de esos algoritmos de «ver» la información que pretendemos obtener, de un modo muy similar a como sucede con un observador humano.
- **Extracción de información**. Los valores de las distintas bandas de una imagen pueden emplearse para derivar otros nuevos, tales como variables de tipo físico o pertenencia a clases predefinidas que pueden identificar los elementos presentes sobre el terreno. Estos procesos implican una *interpretación* de la imagen en cuestión.

17.1.3. Análisis visual y análisis automático

Las tareas de corrección y mejora van a implicar necesariamente la utilización de algoritmos, los cuales veremos en las siguientes secciones. Estas operaciones, como hemos dicho, favorecen el buen desarrollo posterior de la extracción de información, y mejoran los resultados obtenidos. Las operaciones de extracción de información, al contrario que las anteriores, pueden llevarse a cabo tanto manualmente y de forma visual(por ejemplo, digitalizando directamente en pantalla), como de forma automatizada (veremos algoritmos para ello en este capítulo, así como en otros posteriores). Existe, pues, un análisis manual que requiere un operario visualizando y tratando una imagen, y otro automático en el que se deja al propio SIG realizar esta tarea.

Algunas de las principales diferencias entre el análisis visual y el automático se enuncian a continuación. Estas condicionan la utilización de una u otra solución en función de las necesidades.

- El análisis visual requiere conocimientos más sencillos e intuitivos, mientras que el automático requiere conocimientos técnicos para ajustar los algoritmos que realizan el proceso.
- Mientras que el análisis visual se basa tanto en las propiedades espaciales como en las tonalidades de la imagen, en el caso de un análisis automático, este se basa fundamentalmente en los valores de reflectancia, incorporando en ocasiones cierta componente espacial. Las propiedades de forma, tamaño u orientación, aunque importantes como vimos en el apartado anterior, resultan más difíciles de ser reconocidas y empleadas por los algoritmos correspondientes que las basadas en el análisis de los Niveles Digitales.
- El ojo humano puede analizar imágenes en blanco y negro o en color, pero es difícil que pueda trabajar con imágenes con más bandas, que pueden llegar incluso a cientos. Por ello, las imágenes multi e hiperespectrales se analizan de forma habitual mediante algoritmos y procesos automatizados, mientras que el análisis visual queda limitado para las imágenes más sencillas. Como veremos más adelante, pueden crearse representaciones en color de las imágenes con

elevado número de bandas, acercando de ese modo la posibilidad de un análisis visual de estas, aunque en ningún caso de la totalidad de bandas, sino solo de un subconjunto muy reducido de ellas. De igual modo, una banda aislada puede analizarse visualmente, pero ello no permite acceder a toda la información que el conjunto de bandas de la imagen puede aportar, mientras que el análisis no visual no se encuentra restringido por las limitaciones propias de la visión humana.

- La naturaleza del análisis automatizado es fundamentalmente cuantitativa, incorporando formulaciones estadísticas o análisis matemáticos basados en los valores que se recogen en cada píxel. Por el contrario, el análisis visual se basa en propiedades cualitativas de los elementos de la imagen, más sencillas de percibir e interpretar por quién desarrolla dicho análisis.

Es importante tener en cuenta estas diferencias, ya que las circunstancias en las que va a ser necesaria la interpretación de una imagen pueden ser muy variadas, debiendo en cada caso procederse a ella de la forma más adecuada.

17.2. Correcciones y preparación de imágenes

Del mismo modo que un MDE contiene depresiones irreales (véase capítulo 16) que deben ser corregidas antes de proceder a su análisis hidrológico, las imágenes presentan también otros elementos que hacen necesaria una preparación antes de pasar al uso propiamente dicho. Estos elementos están relacionados con la forma en que los sensores capturan la información y las imprecisiones que pueden aparecer en este proceso, y tratan de eliminar estas imprecisiones para que los Niveles Digitales de las distintas celdas reflejen fielmente la realidad presente en el momento de capturar dicha imagen.

Independientemente del número de bandas de la imagen, las correcciones se realizan de forma individual para cada una de las bandas, ya que algunos de los errores pueden aparecer solo en parte ellas, y no necesariamente sobre todas de forma simultanea para un mismo píxel.

Los tipos principales de deficiencias que puede contener una imagen son los siguientes:

- **Geométricas**. Distorsiones de forma provocadas por los movimientos y oscilaciones del sensor, o por el relieve del terreno, entre otros factores.
  Con una base conceptual relacionada con las técnicas de corrección de estas deficiencias, existe un paso básico que debemos llevar a cabo en muchos casos: la georreferenciación de la imagen. Para incorporar dicha imagen dentro del ámbito de un SIG, es necesario que le asignemos una referencia geográfica ya que, como bien sabemos, todos los datos espaciales han de tener definida su posición en el espacio. Cuando esta referencia no existe, es necesario llevar a cabo un proceso que la defina.
- **Radiométricas**. Valores incorrectamente recogidos por el sensor. Pueden ser debidas a errores del sensor o introducidos en el proceso de almacenamiento o transmisión, o bien a distorsiones debidas al efecto que la atmósfera ejerce sobre la radiación (y por tanto sobre los valores recogidos).

17.2.1. Corrección geométrica

Por la propia naturaleza del proceso de toma de imágenes, estas presentan habitualmente diversas formas de distorsión geométrica. Entre ellas, cabe mencionar las debidas a los siguientes factores [224]

- La rotación de la Tierra durante la toma de la imagen.
- La frecuencia de escaneo del sensor.

- La curvatura de la Tierra.
- Los efectos panorámicos.
- Las variaciones en la velocidad y altura de la plataforma de registro.

Para disminuir estos efectos, existen una serie de métodos cuya finalidad es corregir y distorsionar la imagen original con objeto de que esta constituya una representación más fiable de la escena original [43]. Dos son los métodos más utilizados para la corrección geométrica de las imágenes: la *rectificación* y la *ortorectificación*.

Rectificación

El proceso de rectificación se fundamenta en el establecimiento de una correspondencia entre las coordenadas de los píxeles de la imagen y las coordenadas reales sobre el terreno de los objetos que estos píxeles representan. Por medio de una función de transformación se establece una correspondencia que permite modificar la imagen original y obtener una nueva, la cual refleja más fielmente el terreno y corrige los errores geométricos que han aparecido durante la toma de la imagen.

El proceso de rectificación implica el cálculo de la función de transformación, para lo cual se emplea un conjunto de *puntos de referencia* o *puntos de control*[1]. Estos son elementos puntuales de los cuales se conoce con exactitud sus coordenadas reales (bien habiéndolas tomado sobre el terreno o bien con una imagen ya corregida cuya georreferencia pueda emplearse como dato fiable), y pueden localizarse en la imagen. Identificando estos puntos y localizándolos sobre la imagen a corregir se puede calcular la transformación de tal forma que al aplicar dicha transformación sobre los puntos de control, pasen a tener las coordenadas que realmente les corresponden con un error mínimo.

Los puntos de control deben buscarse en elementos del terreno que sean fácilmente localizables y que puedan definirse con la mayor precisión posible. Cruces de caminos u otros puntos singulares se emplean habitualmente para este fin. Lógicamente, la elección de estos puntos depende de la resolución de la imagen, ya que elementos tales como vértices geodésicos, muy aptos para ser puntos de control, pueden verse en las imágenes de gran resolución pero no en las de menor detalle.

Con los puntos de control determinados, debe elegirse un tipo de función para ajustar. Una transformación afín puede ser una opción inicial, pero las modificaciones que permite (rotación, traslación y cambio de escala) no son suficientes para la corrección de las distorsiones geométricas habituales, ya que estas son de una naturaleza más compleja. Funciones de segundo o tercer grado son las opciones más habituales que cumplan los requisitos necesarios, ajustándose estas mediante mínimos cuadrados.

El número de puntos de control debe ser mayor al estrictamente necesario para poder realizar el ajuste, ya que la redundancia que aportan los puntos adicionales da significación estadística a la bondad de este.

En base a los puntos de control puede calcularse la bondad del ajuste mediante el cálculo del error medio cuadrático. Por muy precisa que sea la transformación que realicemos, esta nunca va a asignar a todos los puntos de control la coordenada correcta que les corresponde, y que es la que hemos usado para definir dicha transformación. Las discrepancias entre esas coordenadas reales de los puntos de control y las asignadas a los mismos a partir del polinomio ajustado defi-

[1]Estos puntos de control son similares a los que veíamos en el caso de utilizar una tableta digitalizadora, cuyo uso explicamos en el capítulo 6

nen el error medio cuadrático del ajuste según la siguiente expresión:

$$\mathrm{EMC} = \sqrt{\frac{\sum_{i=1}^{N}(x_i - x'_i)^2 + (y_i - y'_i)^2}{N}} \tag{17.2.1}$$

Para poder calcular el error medio cuadrático, se necesita al menos un punto de control más que el mínimo necesario para ajustar la función polinómica utilizada.

El proceso de rectificación no solo puede emplearse para la corrección geométrica de una imagen que dispone de una referencia geográfica incorrecta por presentarse distorsiones y deformaciones. También puede utilizarse para georreferenciar imágenes que carecen por completo de toda referencia geográfica.

Este caso, como vimos, es el habitual cuando utilizamos imágenes aéreas en formato impreso o mapas en papel, que han de ser escaneados (digitalizados) para poder incorporarse como datos dentro de un SIG, según vimos en el apartado 6.5.2. El proceso de escaneado digitaliza la imagen pero no la referencia geográfica de esta en caso de que exista. Las marcas fiduciales o las coordenadas señaladas en un mapa no son interpretadas como tales una vez la imagen resultante de la digitalización se incorporar a un SIG. Por ello, es necesario definir dicha referencia geográfica para que la imagen adquiera todo su significado.

El proceso se lleva a cabo de igual modo, localizando puntos de control en la imagen (de los cuales solo se conocen sus coordenadas dentro de la imagen, en el espacio de referencia de los píxeles), y dándoles coordenadas reales en base a las cuales pueda definirse la transformación correspondiente.

La imagen 17.1 muestra un ejemplo gráfico de este proceso. Se dispone de un mapa escaneado sin coordenadas, y una imagen de la que sí se conocen las coordenadas de cada píxel, pues ha sufrido un proceso de georreferenciación y corrección previo. Localizando un número suficiente de puntos singulares en ambas imágenes, puede establecerse la transformación que permite asignar coordenadas a todos los píxeles del mapa escaneado. Una vez el proceso se ha realizado, este mapa escaneado puede ya emplearse dentro de un SIG junto con cualquier otra capa que represente esa misma porción del terreno, pues contienen toda la referencia geográfica necesaria.

Ortorrectificación

La ortorrectificación persigue un objetivo similar a la rectificación, pero incluye un factor adicional: la elevación. En lugar de darse una transformación plana, esta es de tipo espacial, ya que utiliza los valores de elevación correspondientes a los distintos píxeles para proceder al ajuste.

Esta elevación provoca un desplazamiento aparente de los píxeles y, según el caso, puede ser tan necesario corregir este efecto como lo es el hacerlo para las distorsiones del plano. En la figura 17.2 se muestra una fotografía aérea antes y después del proceso de ortorrectificacion. A esta segunda se la denomina *ortofotografía*.

La figura 17.3 muestra claramente el efecto de las distorsiones que derivan del relieve del terreno o de la propia altura de los elementos fotografiados, tales como grandes edificios.

La ortofotografía tiene una escala uniforme, y es como una versión fotográfica de un mapa de esa zona. Por esta razón, pueden medirse áreas y distancias en ella, o utilizarla para superponer sobre ella otras capas. En una imagen sin ortorrectificar, y cuando las distorsiones geométricas derivadas del relieve son notables, la medición daría lugar a valores erróneos y la superposición con otras capas no tendría una coincidencia completa.

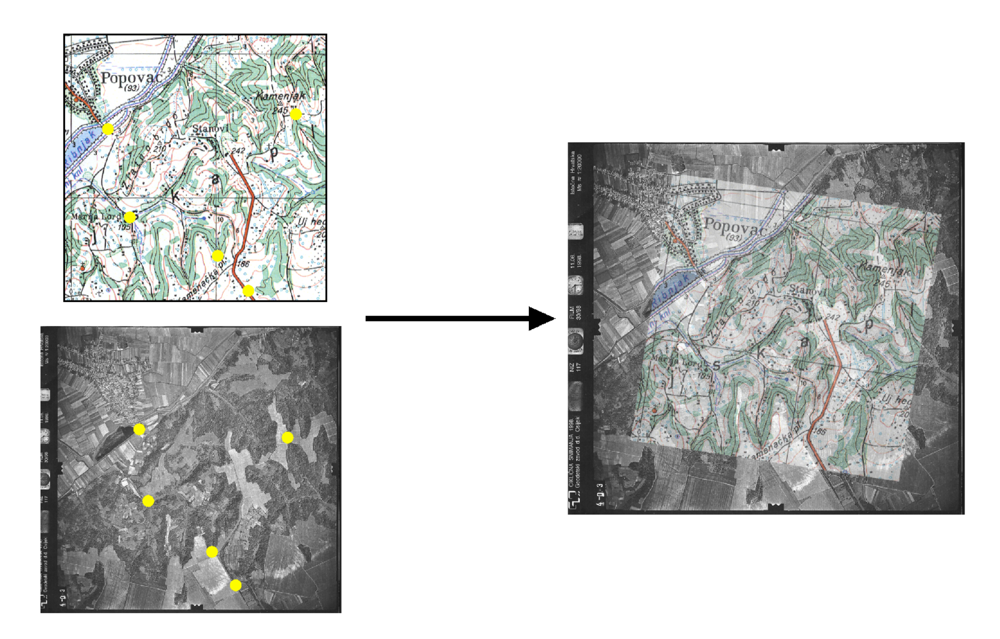


Figura 17.1: Proceso de referenciación de una imagen (mapa escaneado), empleando para la localización de puntos de control (en amarillo) una fotografía aérea ya georreferenciada.

Puesto que se emplean las elevaciones, el concurso de un Modelo Digital de Elevación es imprescindible, pudiendo este proceder del mismo origen que la propia imagen a ortorectificar (en caso de que el sensor ofrezca igualmente pares de imágenes estereoscópicas, y procediendo de la forma que vimos en la sección 6.5.4), o bien de otra fuente distinta.

El proceso de ortorrectificación es complejo y tiene una base matemática extensa. El lector interesado puede encontrar más detalles, por ejemplo, en [225].

17.2.2. Corrección radiométrica

La corrección radiométrica corrige valores erróneamente registrados o ruido presente en la imagen, entre otras alteraciones que pueden aparecer en esta. Existen muchas razones por las que el Nivel Digital recogido en un píxel puede no ser correcto y no reflejar con fidelidad la realidad presente sobre el terreno.

En algunos casos el valor que aparece en el píxel es un valor erróneo del cual no puede extraerse información alguna. Este caso se presenta, por ejemplo, si en el proceso de almacenamiento se ha producido una mala transmisión y algunos píxeles se han «perdido». La información necesaria para establecer un valor correcto en dichos píxeles no puede tomarse del propio píxel, sino que es necesario recurrir a los píxeles circundantes o a otros elementos tales como el histograma de la imagen.

En otros casos, el Nivel Digital del píxel sí contiene información, y lo que es necesario es transformar este, ya que, debido a la influencia de factores externos, no es exactamente el que debiera corresponderle. El conocimiento de esos factores externos es necesario para poder realizar esta transformación.

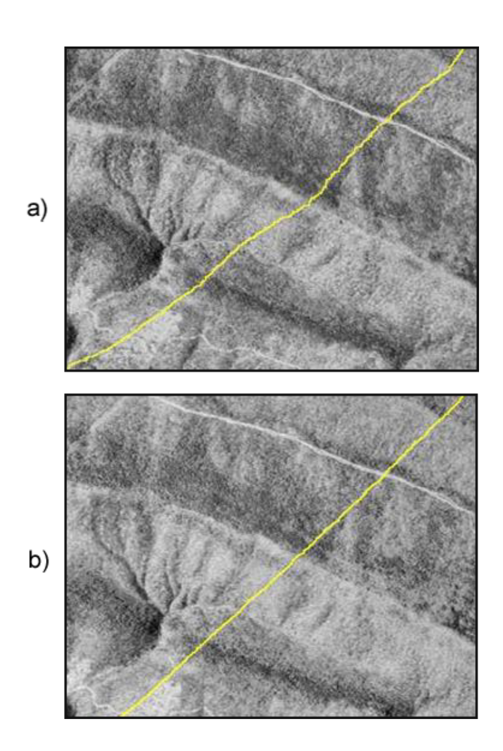


Figura 17.2: Fotografía aérea antes y después del proceso de ortorrectificación

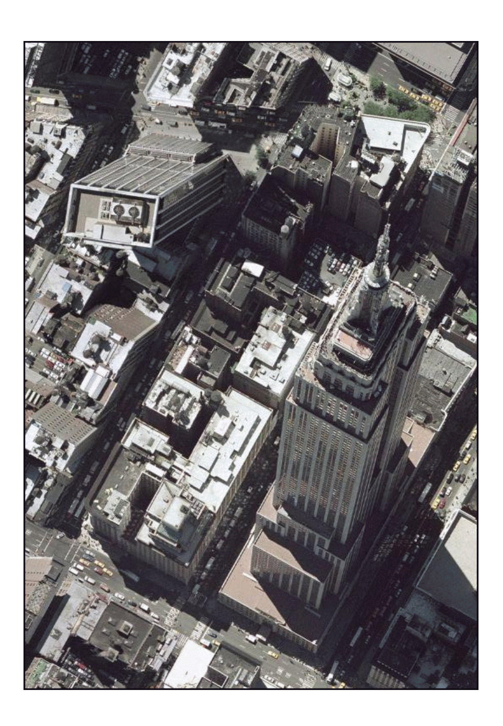


Figura 17.3: Las distorsiones provocadas por el hecho de que la escena fotografiada no es plana se aprecian con claridad en esta imagen.

En función de la causa que los ha originado, distinguimos los siguientes tipos de errores:

- **Errores debidos a los sensores**. Los sensores pueden introducir errores en las imágenes tanto en forma de distorsión como en forma de ruido, ya sea este regular o aleatorio.
  Por ejemplo, los sensores ópticos pueden presentar distorsiones achacables a las lentes, que con frecuencia se manifiesta en áreas más oscuras en el borde de las imágenes en comparación con las celdas centrales. En el caso de sensores electro-ópticos, deben emplearse datos auxiliares para la calibración de estos y así garantizar el registro correcto de los valores correspondientes.
  Errores habituales dentro de este grupo son igualmente los píxeles o líneas perdidas, apareciendo píxeles aislados o líneas completas cuyos valores deben descartarse. La estimación de valores para dichos píxeles o líneas se realiza empleando los valores de píxeles circundantes, pues existe una relación clara entre ellos. El empleo de filtros (ver más adelante el apartado 17.3.2), es una técnica habitual para realizar esta corrección.
  Existe también correlación entre las distintas bandas de una imagen, por lo que no solo pueden utilizarse los píxeles de la misma banda, sino también los propios píxeles perdidos, pero en otras bandas. En general, los sensores que recogen las distintas longitudes de onda (las distintas bandas) son independientes, por lo que el error no debe aparecer en todas ellas.
- **Errores debidos a la topografía**. Además de producir distorsiones geométricas como ya hemos visto, las formas del relieve condicionan la forma en que la radiación se refleja, ya que di-

cha reflexión esta condicionada por el ángulo de incidencia. El uso de Modelos Digitales de Elevaciones e información sobre las condiciones de iluminación en las que se ha tomado la imagen permite plantear modelos de iluminación que pueden corregir estos efectos.

Existen diversas formas de aplicar una corrección a una imagen y eliminar el efecto de la topografía, algunas de ellas relativamente simples. Una formulación simple es la siguiente:

- A partir de un Modelo Digital del Elevaciones se calcula una capa de relieve sombreado. Los parámetros empleados para su creación (azimut y elevación) deben coincidir en la medida de lo posible con las condiciones existentes en el momento en que la imagen fue recogida.
- Se realiza una regresión entre esta capa de relieve sombreado y la imagen a corregir, de forma que se tenga una función de la forma $y = ax + b$.
- Se aplica la transformación definida por la anterior ecuación a los valores de la capa de relieve sombreado.
- Se resta la capa resultante a la imagen a corregir.
- Se suma a la capa resultante del paso anterior el valor de reflectancia media de la imagen original.

▪ **Errores debidos al efecto de la atmósfera en la radiación**. Los errores debidos a la atmósfera son de los más importantes dentro de los que provocan alteraciones radiométricas en la imagen, y se estudian en ocasiones de forma independiente, en lugar de como un subtipo de error radiométrico.

Para comprender la necesidad de esta corrección debe tenerse en cuenta que en algunos casos lo que interesa de la imagen no son los Niveles Digitales, sino una variable con sentido físico como la radiancia correspondiente a cada longitud de onda. Será esta radiancia la que luego se analice para la obtención de otros parámetros físicos derivados, y su obtención se realiza a partir de los Niveles Digitales aplicando ecuaciones lineales con parámetros dependientes del sensor.

El problema estriba en que la radiancia que se obtiene al aplicar este proceso es la que ha alcanzado el sensor, que no ha de corresponderse necesariamente con la que se tiene sobre el terreno o la que recibiría el sensor si estuviera a una distancia mínima del objeto. La atmósfera afecta a la radiación en su camino desde el suelo hasta el sensor, y distorsiona la información recogida.

Los efectos atmosféricos son principalmente de dos clases: *difusión* y *absorción*. La difusión es causada por las partículas de pequeño tamaño de la atmósfera, que desvían una parte de la energía radiante, alterando su dirección. La modificación que se produce depende del tamaño de las partículas implicadas en relación con la longitud de onda de la radiación [226]. La absorción, por su parte, se produce cuando los elementos constituyentes de la atmósfera absorben parte de la radiación para emitirla posteriormente en una longitud de onda distinta. La intensidad de la radiación disminuye con el efecto de la difusión. Ambos efectos conjuntos producen un efecto de «neblina» en la imagen, restándo contraste. La difusión, asimismo, tiene un efecto adicional de adyacencia, ya que cada píxel recoge parcialmente la radiación que en realidad debería corresponder a otros píxeles contiguos.

La corrección de los efectos atmosféricos es compleja y existen muchas formulaciones distintas que escapan al alcance de este texto. Para saber más, una buena descripción de estos métodos puede encontrarse en [227]

17.3. Mejoras

Una vez las imágenes están corregidas, puede procederse a su análisis. Este, en especial el de tipo visual, puede proporcionar mejores resultados si se tratan los valores de la imagen con anterioridad, efectuando ajustes encaminados a mejorar las imágenes como datos de partida.

Podemos comprender mejor esto si pensamos en el tratamiento de imágenes fotográficas habituales. Los métodos que se aplican para realzar el color de estas, enfocar zonas borrosas, o ajustar el brillo o el contraste, entre otros, se pueden aplicar de idéntica forma a imágenes aéreas o de satélite. El efecto que tienen sobre estas es similar, mejorando la forma en que las percibimos, y también la forma en que sus valores pueden ser analizados en otros procesos posteriores.

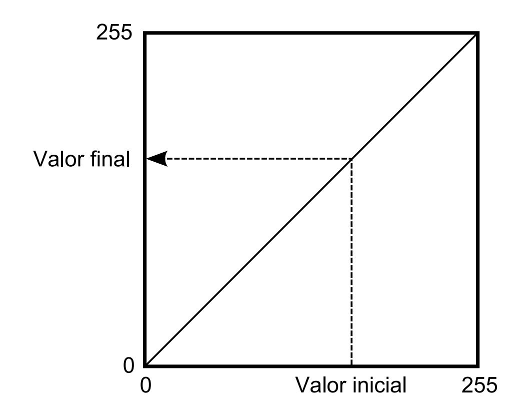
Todos estos procesos son elementos del análisis de imágenes digitales genéricas, que al ser aplicados sobre imágenes aéreas o de satélite mantiene su mismo significado y utilidad a pesar del contexto particular en el que se utilizan.

17.3.1. Operaciones píxel a píxel

Las operaciones de esta clase se corresponden con las funciones de tipo local del álgebra de mapas, ya que toman el Nivel Digital de un píxel de la imagen original y únicamente en función de su valor calculan el del mismo píxel en la imagen resultante. Por ello, todas estas operaciones tienen la forma

$$ND' = f(ND) \qquad (17.3.1)$$

Según sea la función, tenemos un tipo u otro de operación, con un efecto concreto. Para mostrar estas funciones de forma gráfica, emplearemos gráficas como la siguiente, la cual se corresponde con la transformación identidad.



La gráfica establece una relación entre el Nivel Digital original y su tonalidad asociada (en abscisas) y los correspondientes en la imagen resultante (en ordenadas). Aplicando esta transformación a todos los píxeles, se obtiene la imagen mejorada.

A la hora de definir estas funciones de transformación, un elemento de utilidad es el histograma, ya que da una visión global de los Niveles Digitales presentes en el conjunto de píxeles de la imagen. La distribución de Niveles Digitales representada en el histograma aporta una información a partir de la cual pueden definirse estas operaciones de mejora a nivel de píxel.

Segmentación

La primera operación que veremos es la *segmentación*, cuyo objetivo es particionar una imagen en diversas regiones en base a criterios de homogeneidad o heterogeneidad [228].

Aunque veremos que existen muchas formas de segmentación, en su forma más básica se trata de una operación píxel a píxel cuyo resultado es lo que conocemos como una imagen *binaria*. Esto quiere decir que dicha imagen tiene únicamente dos valores: uno de ellos representa lo que se conoce como *objetos* (que se representan habitualmente en negro), y el otro representan el *fondo* de la imagen (habitualmente blanco).

Este procedimiento nos sirve para separar partes de la imagen en función de sus valores. Muy frecuentemente, encontraremos en una imagen algunos elementos que nos resultarán de interés (por ejemplo, una carretera a lo largo de una zona desierta, claramente diferenciable), y resultará de interés separar esos elementos del resto de la imagen. Esto permitirá que sea más sencillo operar con esos elementos en posteriores operaciones. La segmentación simplifica los elementos de la imagen, y la imagen resultante tiene menor complejidad pero resulta más útil a la hora de realizar tales operaciones.

La figura 17.4 nos muestra el resultado de una operación de segmentación aplicada sobre una imagen base (en este caso se trata de una ortofoto en blanco y negro). En los sucesivos apartados se empleará esta misma imagen original para ilustrar cada procedimiento. La curva de transformación asociada, con un umbral aplicado de 173, se muestra en la figura 17.5.

Como vemos, la función de transformación simplemente establece un valor umbral, a partir del cual se asocia el valor 255, mientras que por debajo se asocia el valor 0 (es habitual también emplear los valores 1 y 0 en lugar de los anteriores). Suponemos en este caso que los valores que resultan de interés y queremos aislar son aquellos que presentan valores superiores al umbral, mientras que todos los restantes los consideramos parte del fondo. La operación puede realizarse también al contrario, en caso de que los objetos a separar vengan caracterizados por valores bajos. En este caso, se asignaría 255 a los valores por debajo del umbral y 0 a los restantes.

La selección del umbral es el elemento clave en esta forma de segmentación, y condiciona el resultado del procedimiento. El umbral puede seleccionarse en base a las propias características de la imagen, o bien a partir de un conocimiento previo de las características de los objetos a definir. En el primer caso, el histograma de la imagen se convierte en una herramienta base para definir el umbral correctamente.

Además de establecerlo de forma directa mediante un valor concreto, existen diversas formulaciones que permiten fijar un umbral de forma automática, siendo formulaciones adaptativas. De igual modo, este umbral puede no ser constante para toda la imagen, y realizarse una segmentación por zonas, de tal modo que no se emplee el mismo umbral en todas ellas. Esto puede mejorar la eficacia del proceso, ya que la relación entre el fondo y los objetos a separar no ha de ser la misma en todas las áreas de la imagen.

En [229] se describen formulaciones alternativas para esta definición automática de umbrales.

Este tipo de segmentación es, como se ha dicho, la forma más sencilla de esta clase de procesos, ya que la imagen de partida es una imagen en escala de grises (imagen de una banda), y el resultado es una imagen binaria. Se pueden, no obstante, segmentar imágenes multibanda y hacerlo no en dos partes (fondo y elemento), sino en un número mayor de ellos (fondo y distintos elementos, codificados con distintos valores), entrañando las formulaciones correspondientes una complejidad lógicamente mayor. Existen, por tanto, otros métodos de segmentación, muchos de los cuales no implican únicamente operaciones píxel a píxel, sino más complejas. Aunque no entraremos a detallarlas con tanta profundidad, se enumeran y describen brevemente a continuación las principales familias en que podemos dividir dichos métodos [230]:

- **Segmentación basada en umbral**. Del tipo de los que hemos visto hasta el momento.
- **Segmentación basada en regiones**. Se ha de proporcionar al algoritmo una serie de puntos que definan las distintas regiones, tales como puntos centrales de las mismas. A partir de ellos, se va ampliando el espacio ocupado por cada

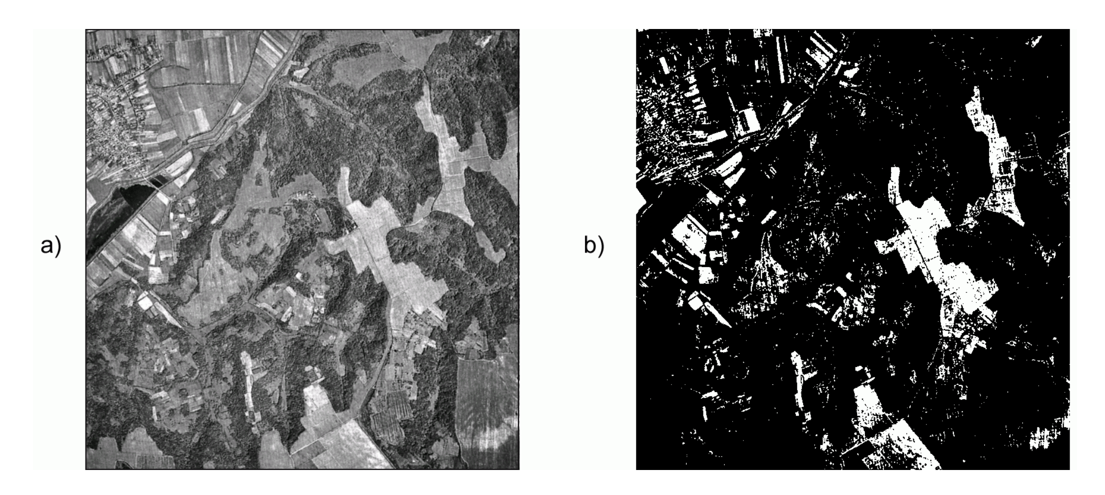


Figura 17.4: Imagen original (a) y resultado de un proceso de segmentación (b)

región (de un único punto) en función de la homogeneidad de las celdas circundantes[2].

- **Segmentación basada en agregación**. Métodos iterativos que agrupan los píxeles en grupos de forma que estas agrupaciones sean lo más homogéneas posibles. Estos algoritmos no los vamos a tratar en esta sección, pero sí que se describen en otro punto de este texto, dentro de la sección 21.2.2 dedicada a la clasificación no supervisada.

Para el lector interesado, en [231] o [228] pueden encontrarse tratados en profundidad todos estos distintos grupos de metodologías.

Desde el punto de vista de las operaciones del álgebra de mapas que conocemos, la segmentación es una reclasificación de valores lo más sencilla posible, ya que únicamente se crean dos clases (la de valores por debajo del umbral y la de valores por encima de él). Un proceso algo más complejo es la definición de varios umbrales, de forma que aparezcan más clases, siendo este también equivalente a la reclasificación según la vimos en su momento. En el ámbito del análisis de imágenes esta operación se conoce habitualmente con el nombre de *Density slicing*.

Expansión de contraste

El proceso de expansión de contraste (también conocido como *normalización*) aporta mayor definición a los distintos elementos de la imagen. Esto es así porque la expansión del contraste modifica el histograma de tal forma que se utilicen todos los valores posibles de los Niveles Digitales. Es decir, que de los 255 niveles de intensidad de gris el histograma abarque todos ellos, y al traducirse eso en las tonalidades de gris a la hora de representarse, estas también vayan desde el valor 0 (negro) hasta el 255(blanco). Logrando esto, se obtiene una mayor definición visual que puede apreciarse en la imagen.

[2]Es probable que el lector haya utilizado alguna vez algún programa de tratamiento de imágenes, en cuyo caso podrá entender de forma sencilla la forma de proceder de estos algoritmos, asimilándolo a la herramienta «varita mágica», que selecciona los píxeles contiguos y de un color similar a uno seleccionado sobre la imagen. En este caso, los puntos originales se le proporcionan al algoritmo, que a partir de ellos efectúa un procedimiento similar al de dicha «varita mágica», seleccionando progresivamente los píxeles «parecidos» situados alrededor, haciendo crecer cada región.

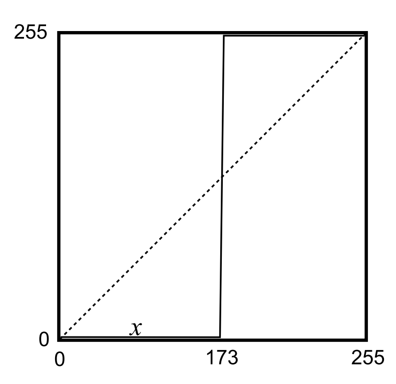


Figura 17.5: Curva de transformación asociada a una segmentación por umbral.

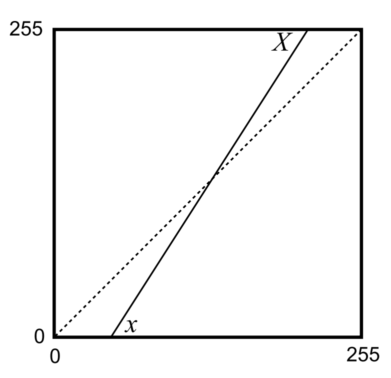


Figura 17.6: Curva de transformación correspondiente a una expansión de contraste

Este efecto se consigue escalando el histograma, «estirándolo» horizontalmente para que cubra todo el rango de valores. En imágenes cuyos Niveles Digitales se sitúan mayoritariamente en un intervalo reducido, la representación presenta tonalidades homogéneas que no aprovechan todo el contraste que puede lograrse con el rango completo de intensidades. Aplicando una transformación conveniente, se consigue que el histograma gane amplitud y la imagen gane en contraste.

A partir de una imagen con sus valores mínimo (x) y máximo(X) respectivos, se trata de obtener una imagen que mantenga una distribución similar en su histograma, pero de tal forma que los valores mínimo y máximo pasen a ser otros distintos, en particular 0 y 255 como valores límites del rango habitual de Niveles Digitales. Para lograr esto, la forma más simple es aplicar una fórmula como la siguiente.

$$ND' = \frac{255(ND - x)}{X - x} \qquad (17.3.2)$$

donde ND es el Nivel Digital de la imagen original y ND' el de la imagen resultante.

Si recordamos lo visto en el apartado 15.2.1, la normalización era una función de tipo local dentro del álgebra de mapas, y respondía a una expresión similar a la anterior. El fundamento en este caso es el mismo, aunque la finalidad sea aquí la de mejorar la imagen de cara a su visualización.

En el tratamiento de imágenes es habitual no aplicar directamente la expresión 17.3.2 con los valores mínimo y máximo absolutos de la imagen. De hacerlo así, la presencia de valores atípicos alejados de la media (píxeles muy o muy poco brillantes) puede desvirtuar el rango en el que se sitúan realmente los Niveles Digitales de la imagen, y no conseguirse mediante la expansión de contraste el efecto deseado.

Una solución es acudir al histograma y tomar como valores de x y X no los extremos absolutos, sino los correspondientes a los percentiles del 5 % y el 95 %. Al operar con estos valores límite, aparecerán valores resultantes fuera del rango de destino (0–255 en el caso más habitual que citamos), que se igualan a los valores mínimo o máximo de dicho rango en función de si se encuentran fuera por el lado negativo o el positivo.

Otra solución frecuente para la selección de x y X consiste en tomar la frecuencia del pico del histograma (es decir, el número del píxeles que presentan el valor máximo), y definir un umbral utilizando un porcentaje de dicha frecuen-

cia. El valor mínimo x corresponde al valor del primer Nivel Digital que, empezando desde 0 y en sentido creciente (de izquierda a derecha del histograma), supera dicha frecuencia. De modo similar, el valor máximo X se calcula recorriendo el histograma en sentido negativo (de derecha a izquierda) y tomando el primer Nivel Digital cuya banda de frecuencia en el histograma sea superior a la establecida por el umbral.

Si se aplica la normalización a una imagen multibanda, es importante tener en cuenta que las relaciones entre los Niveles Digitales de estas han de preservarse. Cuando se aplica normalización sobre un conjunto de capas ráster, el objetivo es escalar las distintas variables para que ocupen un rango homogéneo, y esta relación no es relevante. En el caso de imágenes, y trabajando con bandas en lugar de capas, sí lo es, ya que de ella dependen, por ejemplo, los colores obtenidos al componerlas para formar imágenes en falso color según veremos más adelante.

Modificación de brillo y contraste

Tanto el brillo como el contraste pueden modificarse mediante funciones lineales sencillas, tal como puede verse en la imagen 17.7.

Suponiendo como es habitual que trabajamos en un rango de valores entre 0 y 255, la modificación del brillo se lleva a cabo con una función de la forma

$$ND' = ND + c \qquad ; \qquad c \in [0, 255]. \tag{17.3.3}$$

Valores positivos de la constante c aclaran la imagen, mientras que los negativos la oscurecen.

Para el caso del contraste, las funciones son de la forma

$$ND' = mND + c \tag{17.3.4}$$

El valor de c se calcula a partir del de m, según

$$c = \frac{1-m}{2} \tag{17.3.5}$$

Valores altos de la pendiente y puntos de corte negativos dan lugar a imágenes de alto contraste, mientras que valores bajos de pendiente y puntos de corte positivos producen imágenes de bajo contraste.

El caso de la expansión de contraste es un caso particular de este tipo de transformaciones lineales, que no obstante se ha analizado aparte por sus particularidades y por ser una operación muy frecuente. En concreto, la expansión de contraste es una operación que maximiza el contraste pero sin producir saturación en los tonos extremos (blancos y negros). La función de transformación correspondiente se representa en la figura 17.6.

El empleo de funciones no lineales permite mayor flexibilidad a la hora de mejorar el contraste, al tiempo que se evita la aparición de saturación en los extremos del rango de valores. Una de las funciones más habituales en este caso es la siguiente:

$$ND' = ND^{\alpha} \tag{17.3.6}$$

Esta operación se conoce como *corrección de gamma*, y es habitual en el tratamiento de imágenes.

Las funciones sigmoidales como la mostrada a continuación también son una alternativa frecuente.

$$ND' = \frac{1}{1 + e^{\alpha ND + \beta}} \tag{17.3.7}$$

Ecualización

La ecualización es un proceso similar a la normalización, en cuanto que hace más patentes las

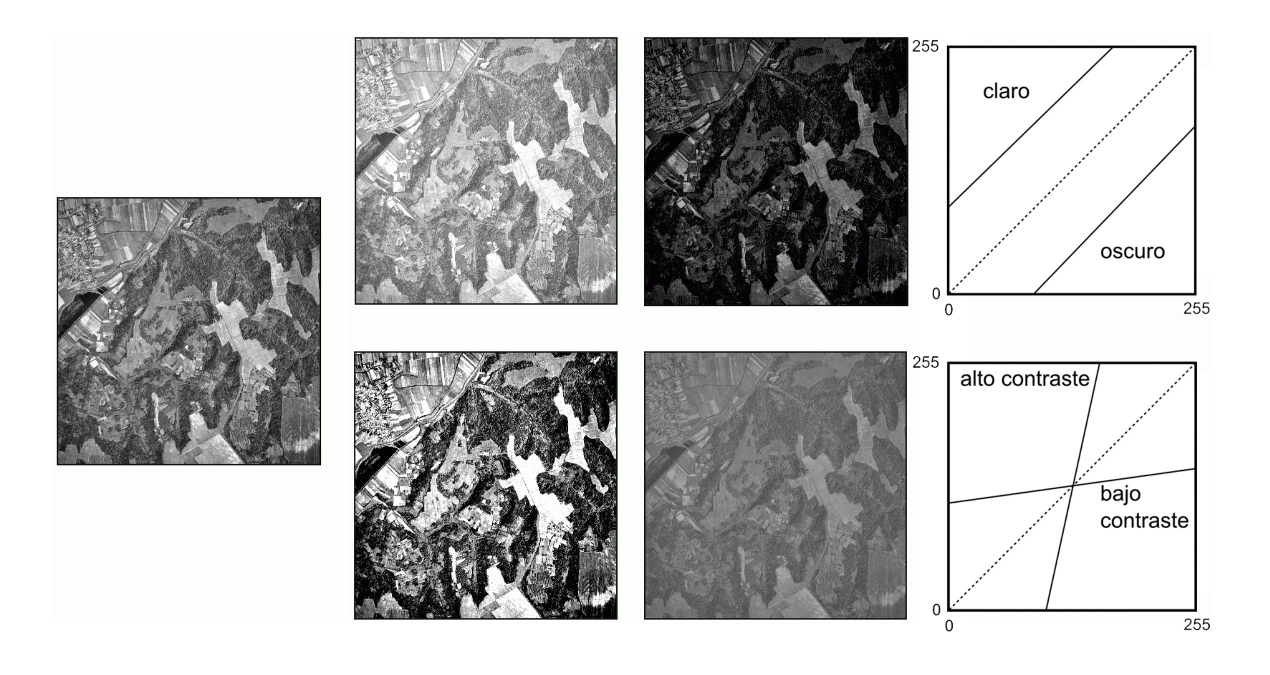


Figura 17.7: Imagen original y modificación del brillo y el contraste, junto con sus curvas de transformación asociadas

diferencias entre los distintos niveles de la imagen, añadiendo definición a esta. El proceso de ecualización modifica el histograma de tal modo que en la nueva imagen todos los niveles digitales tengan la misma frecuencia. De esta forma, se redistribuyen los valores de intensidad a lo largo de todo el espectro, igualando la presencia de estos en la imagen.

La figura 17.8 muestra la comparación entre la imagen original y la ecualizada, así como sus histogramas acumulados. Esto permite ver más claramente la naturaleza de la transformación. Teóricamente, si el histograma es constante tras la ecualización, el histograma acumulado debería ser una recta inclinada. El hecho de trabajar con valores discretos de los Niveles Digitales hace que no sea exactamente así, aunque puede claramente verse que la forma sí se asemeja a una recta en lugar de a la curva sigmoidal que presenta el histograma de la imagen original.

La función de trasformación deriva del propio histograma acumulado original, teniendo su misma forma. De este modo, se hace innecesario el elegir valores umbral sobre el propio histograma, tal como sucedía en el caso de la expansión de contraste. Para un píxel con un Nivel Digital ND en la imagen original, se tiene un valor en la imagen ecualizada dado por la expresión

$$ND' = T(ND)\frac{255}{N} \qquad (17.3.8)$$

siendo N el número total de píxeles de la imagen y $T(x)$ la función que define el histograma acumulado, es decir, el número de píxeles de la imagen con Nivel Digital menor que x.

Al trabajar con imágenes de varias bandas, deben tenerse en cuenta los mismos aspectos ya comentados para la expansión de contraste, en

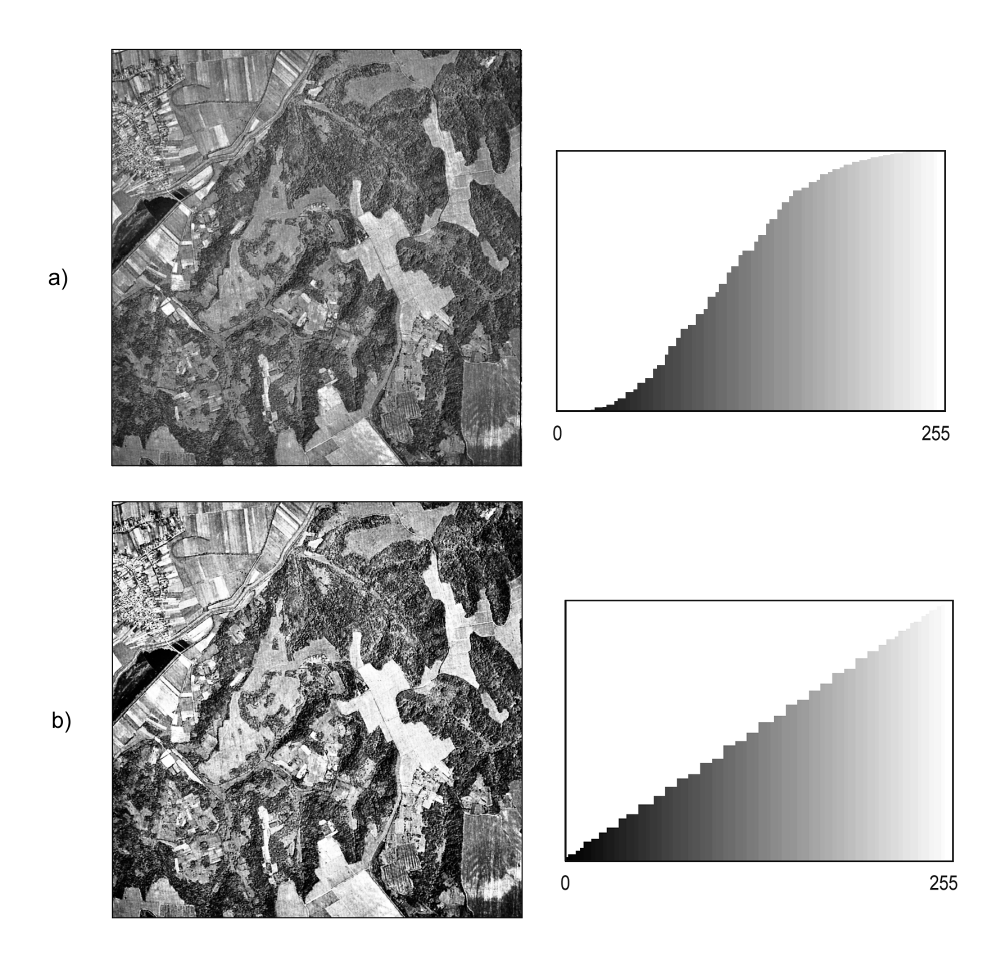


Figura 17.8: a) Imagen e histograma acumulado originales, b) imagen e histograma acumulado tras un proceso de ecualización.

relación a la relación entre las distintas bandas. En [232] pueden encontrarse detalles sobre ecualización de imágenes en color.

17.3.2. Filtros

Al igual que algunas de las operaciones vistas ya en este capítulo, los filtros son similares en concepto a parte de las operaciones que ya conocemos de capítulos anteriores. En particular, comparten sus fundamentos con las funciones focales del álgebra de mapas, aunque, al igual que en otros de los casos anteriores, la utilización de estas tiene aquí una función principalmente relacionada con la mejora visual.

Podemos dividir los filtros en dos tipos en función del tipo de modificación que realizan.

- Filtros de suavizado. Producen un efecto de desenfoque, restando definición a la imagen. Atenúan las diferencias entre píxeles vecinos.
- Filtros de realce: Producen un efecto de enfoque, aumentando la definición de la imagen. Acentúan las diferencias de intensidad entre píxeles vecinos.

Ciertos filtros calculan los nuevos valores como combinaciones lineales de los píxeles del entorno. Son las denominadas *convoluciones*, las cuales ya vimos en el capítulo dedicado al álgebra de mapas. La convolución se define mediante una matriz $n \times n$ (que denominábamos *núcleo*) y esta permite calcular el valor resultante de la operación. En función de los valores de la matriz, el efecto que el filtro tenga sobre la imagen será uno u otro.

Otros filtros, a diferencia de las convoluciones, se basan en funciones estadísticas que no pueden expresarse mediante un núcleo.

Filtros de suavizado

Los filtros de suavizado (también conocidos como filtros *de paso bajo*) provocan una perdida de foco en la imagen. Este efecto se consigue disminuyendo las diferencias entre píxeles contiguos, algo que puede obtenerse por ejemplo mediante un filtro de media. Como ya vimos, este puede expresarse mediante un núcleo como el siguiente:

1	1	1
1	1	1
1	1	1

El efecto del filtro de media vimos que, aplicado sobre una capa de elevación, «redondeaba» el relieve. En el caso de ser aplicado sobre una imagen hace que esta aparezca más borrosa. La cantidad de suavizado puede ser controlada, al igual que en el caso de aplicarse sobre otro tipo de capas ráster, mediante el tamaño de la ventana, que puede ser mayor que el habitual 3×3 anterior.

Otra forma de modificar el efecto del suavizado, en este caso limitándolo, es dando más peso al píxel central. Para ello puede emplearse un núcleo como el mostrado a continuación:

1	1	1
1	2	1
1	1	1

La media es sensible a los valores extremos de la ventana, por lo que una opción habitual es sustituir el filtro de media por uno de mediana. Este no es sensible a la presencia de valores extremos muy alejados de la media (*outliers*), y además garantiza que el valor resultante es un valor que existe como tal en la ventana de píxeles circundantes, lo cual puede resultar de interés en algunas circunstancias.

El filtro de mediana no es una convolución, y no puede por tanto expresarse mediante un núcleo como sucedía con el de media.

La aplicación de un filtro de suavizado no solo tiene como efecto el desenfoque de la imagen, sino que también elimina el ruido de esta. Por ello, estos filtros son una buena herramienta para tratar algunas de las deficiencias de las imágenes, que como vimos pueden presentar píxeles con Niveles Digitales muy distintos de los de su entorno.

En la figura 17.9 puede observarse como la presencia de una línea errónea introducida en una imagen se corrige parcialmente con un filtro de mediana.

Filtros como el de mayoría o moda, que da como valor resultante de la celda aquel que se repite más entre las celdas circundantes, pueden también aplicarse para estos fines.

Las convoluciones tales como el filtro de media no son recomendables en muchos de estos casos, según sea el tipo de ruido que deseamos eliminar. Un tipo de ruido habitual es la presencia de píxeles blancos o negros debidos a errores de

a)



b)

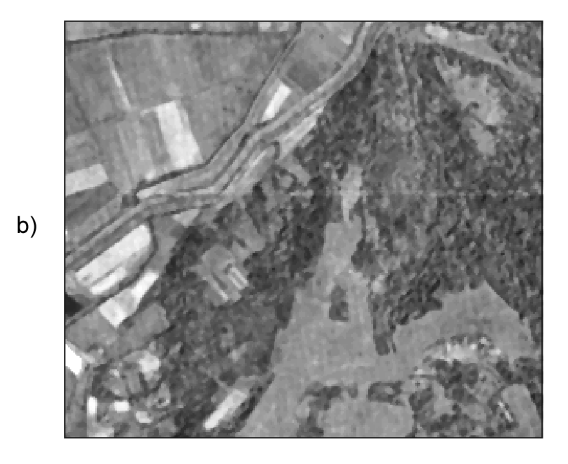


Figura 17.9: Eliminación de una línea errónea por aplicación de un filtro de mediana

transmisión o almacenamiento, es decir, con Niveles Digitales extremos. El sesgo que introducen estos píxeles es muy alto, y la aplicación de un filtro de media no es válido para eliminar esta clase de ruido.

Otro tipo de ruido habitual es el conocido como *sal y pimienta*, que aparece frecuentemente al segmentar imágenes según vimos en un punto anterior. La imagen de salida contiene únicamente valores 1 y 0, representados de la forma habitual en blanco y negro respectivamente. Es normal que en la segmentación aparezcan píxeles con uno de dichos valores, rodeados por pixeles del otro, es decir píxeles completamente aislados. Esto da la sensación de una imagen sobre la que se han «esparcido» píxeles blancos (sal) sobre la parte negra, o negros (pimienta) sobre la parte blanca.

Para limpiar estas imágenes resultantes, se emplean también filtros como los anteriores. Un filtro de mediana aplicado sobre la imagen de la figura 17.4 (resultado de una segmentación) elimina gran parte de los píxeles aislados y genera el resultado mostrado en la figura 17.10.

Filtros de realce

Los filtros de realce (o *de paso alto*) tienen el efecto justamente contrario al de los anteriores, ya que acentúan las diferencias entre píxeles adyacentes, teniendo un efecto de enfoque. La definición aumenta, como puede verse en la imagen 17.11. Si se compara con la imagen original, se aprecia una separación más clara entre las tonalidades de píxeles contiguos, cuya diferencia se acentúa.

Para obtener un filtro de realce, puede partirse de uno de suavizado, ya que una forma de realzar las diferencias entre píxeles es emplear un filtro de paso bajo para obtener una imagen suavizada, y después restar esta imagen de la imagen original. Haciendo esto se elimina el valor medio y se deja en cada píxel el valor que refleja la variación local de los Niveles Digitales.

El núcleo que caracteriza esta transformación puede obtenerse realizando el mismo calculo sobre los núcleos de las operaciones independientes. Es decir, restando al núcleo identidad el de media, obteniendo el siguiente:



Figura 17.10: Eliminación de ruido de tipo *sal y pimienta* mediante filtro de mediana.



Figura 17.11: Imagen tras la aplicación de un filtro de realce

-1	-1	-1
-1	8	-1
-1	-1	-1

Filtros de detección de bordes

Con un fundamento similar a los anteriores, los filtros de detección de bordes permiten localizar las zonas donde se producen transiciones bruscas de intensidad. Aplicado esto al análisis de imágenes en un contexto geográfico, puede emplearse este tipo de filtros como herramientas para identificar zonas y fronteras entre distintas formaciones de una área de estudio.

Un operador habitual para la detección de bordes es el denominado *filtro Laplaciano*, el cual puede expresarse mediante un núcleo de la forma

0	-1	0
-1	4	-1
0	-1	0

Para hacer su expresión más sencilla, tanto este núcleo como los siguientes no se aplican según la ecuación 17.3.9, sino según la siguiente:

$$y = \sum_{i=1}^{n} z_i k_i \; ; \; k_i \in \mathbb{R} \qquad (17.3.9)$$

Es decir, sin dividir el valor resultante entre la suma de los coeficientes.

En la figura 17.12 puede verse el efecto de un filtro Laplaciano tiene sobre dos imágenes. En la imagen a), se ha aplicado sobre la imagen de ejemplo original. En el caso b) se ha empleado como imagen base la imagen ecualizada. Como puede apreciarse, la ecualización realza el efecto del filtro de detección de bordes, haciendo más patentes los resultados. Además de una mejora visual, la ecualización implica en este caso una mejora de los resultados derivados.

Para mostrar más claramente el resultado de estos filtros, se ha escogido una región de la imagen (en la parte superior izquierda) en lugar de la imagen completa, por contener un mayor número de entidades lineales. Como puede apreciarse,

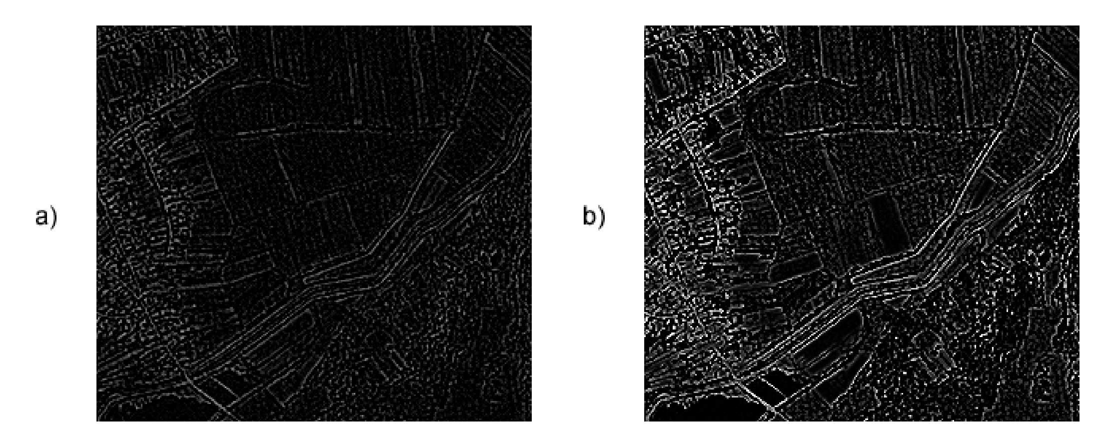


Figura 17.12: Aplicación de un filtro de detección de bordes sobre la imagen original (a) y la imagen ecualizada (b).

es en la detección de estos elementos lineales donde mejor actúan estos filtros.

El filtro Laplaciano tiene su base en el análisis de las segundas derivadas de la función que los Niveles Digitales definen y la detección de puntos donde esta se anule. Otros filtros se basan en el estudio del gradiente (la primera derivada) de forma que se realcen las zonas en las que existan variaciones notables entre las intensidades de píxeles contiguos.

Aplicando filtros de esta segunda clase, la detección de bordes puede hacerse restringida a una dirección predominante, tal como la vertical (detectando variaciones entre filas) o la horizontal (detectando variaciones entre columnas). Por ejemplo, con los filtros siguientes, conocidos como *filtros de Sobel*, cada uno de ellos correspondiente a una de las direcciones anteriores.

1	2	1
0	0	0
-1	-2	-1

-1	0	1
-2	0	2
-1	0	1

El resultado de aplicar estos filtros puede verse en la figura 17.13.

Otros filtros habituales son los siguientes, denominados *filtros de Prewitt*,

1	1	1
0	0	0
-1	-1	-1

-1	0	1
-1	0	1
-1	0	1

Las variantes de estos filtros en las direcciones indicadas pueden combinarse en un filtro global según la expresión

$$y = \sqrt{y_1^2 + y_2^2} \tag{17.3.10}$$

donde y_1 y y_2 son los valores resultantes de la aplicación de los filtros horizontal y vertical respectivamente.

El fundamento de este cálculo es similar a lo visto en 16.5.1 para el caso de la pendiente calculada sobre un MDE.

Con respecto al filtro Laplaciano, los filtros basados en derivadas direccionales tienen como desventaja una mayor sensibilidad al ruido presente en la imagen.

La relación entre los filtros de detección de bordes y los de realce puede verse en el siguiente ejemplo. Si se toma un filtro Laplaciano y se le

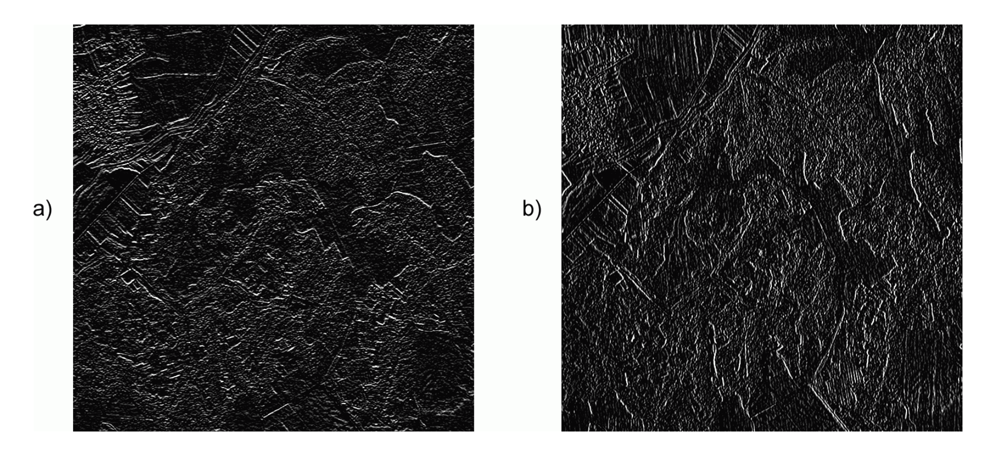


Figura 17.13: Aplicación de un filtro de Sobel vertical (a) y horizontal (b).

añade un filtro identidad (es decir, a la imagen resultante se le suma la imagen original), tenemos el filtro mostrado a continuación.

0	-1	0
-1	5	-1
0	-1	0

Aplicando este filtro se obtiene una imagen que mantiene un aspecto similar a la imagen original, pero con mayor definición. Es decir, el mismo efecto que si se aplicara un filtro de realce.

17.3.3. Fusión de imágenes

La fusión de imágenes engloba a una serie de procesos que permiten integrar la información procedente de varias fuentes de datos distintas en una única imagen. Esta imagen resultante facilita la interpretación y análisis de la información de partida, englobando las características más destacables de las imágenes originales.

Por medio de la fusión pueden crearse imágenes sintéticas que combinen imágenes con alta resolución espacial y otras con alta resolución espectral, y que presenten alta resolución en ambas componentes. Esta operación es muy útil teniendo en cuenta que existen sensores con gran precisión espectral, pero que no se caracterizan por una elevada definición (no dan imágenes con tamaños de píxel bajos), mientras que en otros sucede justamente lo contrario. Unir estas mediante un proceso de fusión es una forma de obtener imágenes de mejor calidad con las que lograr mejores resultados.

El uso combinado de imágenes pancromáticas e imágenes multiespectrales como las del satélite LANDSAT es una práctica habitual a la hora de aplicar este método de fusión.

Veremos tres métodos principales.

- IHS
- Por Componentes Principales
- Brovey

Transformación IHS

La imagen que dispone de una mejor información espectral aporta esta de forma habitual en el modo RGB que ya conocemos. Este, recordemos, implica tres bandas correspondientes a los colores rojo, verde y azul, las cuales dan, por adición, la imagen en color resultante.

El modo RGB es solo uno de los muchos existentes para codificar un color, siendo el más frecuente para el registro de imágenes y la representación de estas en una pantalla, pero no el único. A la hora de llevar a cabo una fusión de imágenes, la utilización de otros formatos resulta más ventajosa.

Uno de los modos que pueden emplearse para este fin es el conocido como IHS[3]. Este modo es más similar a cómo percibimos los distintos colores, y se basa en los tres componentes siguientes[233]:

- **Intensidad (I)**. Expresa el brillo del color. Es la variable que se percibe más sencillamente por parte del ojo humano.
- **Tono (H)**. Expresa de qué color se trata. Más exactamente, indica la longitud de onda predominante de dicho color.
- **Saturación (S)**. Expresa la pureza del color. Valores altos indican una alta presencia de blanco.

Veremos algo más sobre estos conceptos, aunque en un contexto distinto, en el capítulo 27.

Dadas dos imágenes, una de ellas con mayor información espacial y otra con mayor información espectral, puede realizarse una fusión empleando una transformación IHS siguiendo los pasos descritos a continuación [234][235]:

- Se remuestrea la imagen de menor resolución espacial a las dimensiones de la de mayor resolución, o bien ambas a un tamaño de píxel intermedio entre los de ambas imágenes. Ese será el tamaño de píxel de la imagen resultante, mejorando así el de la imagen que aporta la información espectral.
- Se convierten las imágenes al formato IHS. Lo habitual es que la imagen con mayor resolución espacial sea de una única banda, con lo que no es necesaria conversión alguna.
- Se sustituye en la imagen con la información espectral la banda de intensidad I por la banda de intensidad de la otra imagen.
- Se aplica una transformación inversa para pasar de nuevo al formato RGB.

En la figura 17.14 puede verse un esquema del proceso de fusión de imágenes según lo descrito anteriormente.

Transformación por componentes principales

El método de análisis de componentes principales es una técnica estadística cuyos detalles veremos más adelante dentro del apartado 21.5. En esencia, se trata de un método que convierte un conjunto de capas (o bandas) intercorrelacionadas en uno nuevo con un menor número de capas, y menos correlacionadas. Es decir, sustituye las variables originales por otras entre las que no existe correlación, cada una de las cuales explica un porcentaje concreto de la varianza global.

Puede aprovecharse el análisis de componentes principales para efectuar una fusión de imágenes, siguiendo un proceso similar al de la transformación IHS, con los pasos enunciados a continuación.

- Se remuestrean las imágenes a un marco común como ya vimos anteriormente
- Se efectúa un análisis de componentes principales en ambas imágenes.
- El componente principal (la nueva banda que explica un mayor porcentaje) de la imagen con la información espectral, se sustituye por el componente principal de la imagen de mejor resolución espacial.
- Se aplica un proceso inverso al análisis de componentes principales sobre dicha imagen de in-

[3] Intensity, Hue, Saturation (Intensidad, Tono, Saturación)

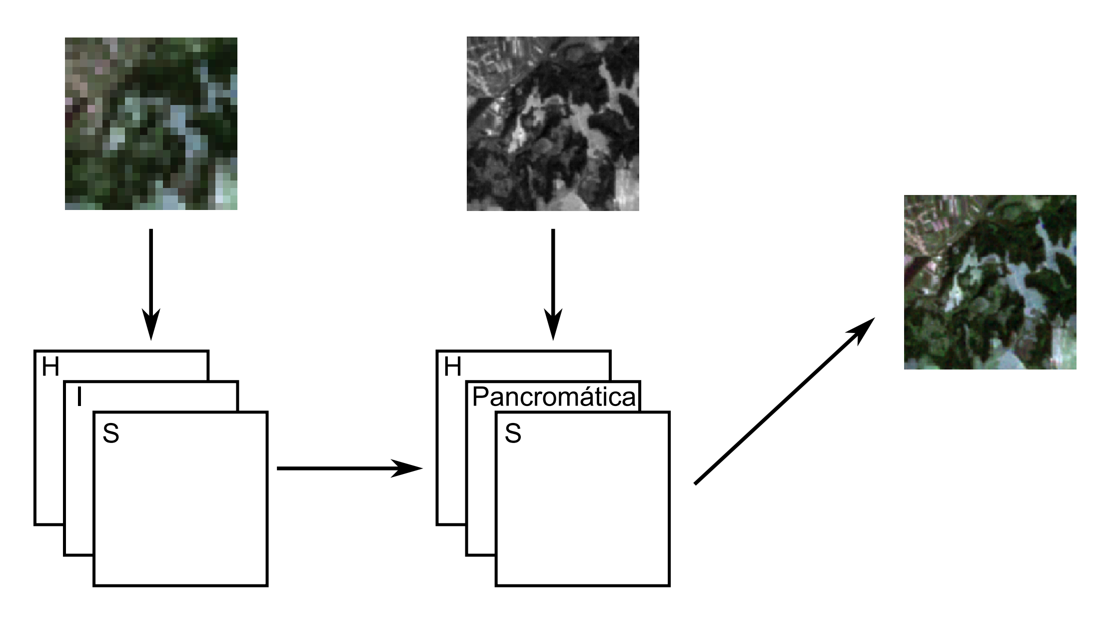


Figura 17.14: Fusión de imágenes mediante transformación IHS

formación espectral, obteniendo así el resultado buscado.

Transformación de Brovey

La transformación de Brovey [236] se emplea también para la fusión de imágenes, aunque su procedimiento es bien distinto a los anteriores. En este caso, no se da una modificación previa del conjunto de capas que conforman la componente multiespectral, seguida esta de una sustitución, sino que se opera banda a banda, aumentando la resolución de estas de forma individual. Combinando las bandas resultantes se obtiene la imagen sintética buscada.

En particular, se parte de una imagen con tres bandas y una imagen pancromática de mayor resolución espacial. Para transformar la banda i-ésima se utiliza la siguiente expresión:

$$ND' = \frac{ND_i}{ND_1 + ND_2 + ND_3} ND_{pan} \quad (17.3.11)$$

donde ND_n es el Nivel Digital del píxel en la banda n-ésima, y ND_{pan} el de dicho píxel en la imagen pancromática. Obviamente, se requieren remuestreos para ajustar las capas a un marco común donde exista correspondencia de píxeles entre ambas imágenes.

Para el caso de los sensores más habituales existen recomendaciones sobre qué tres bandas tomar para definir la imagen multiespectral requerida por el método. Por ejemplo, para el sensor LANDSAT se recomiendan las bandas 2, 4 y 5; y para Quickbird las bandas del verde, rojo e infrarrojo cercano.

17.4. Análisis y extracción de información

Con las imágenes ya preparadas y tratadas con objeto de «exponer» en la mayor medida posible la información que contienen, puede procederse a su análisis. El fin de este análisis puede ser la delineación e identificación de ciertos

elementos (masas de agua, formaciones rocosas, construcciones, etc.), o bien la caracterización de las distintas áreas del terreno (tipo de uso de suelo, condiciones físico–químicas, etc.), existiendo para cada uno de estos propósitos una serie de algoritmos y formulaciones.

17.4.1. Visualización

La obtención de de resultados a partir de imágenes, tales como la clasificación de las distintas zonas o la localización de áreas en las que aparece algún elemento particular del terreno (por ejemplo, zonas de agua como ríos o embalses), se realiza, como veremos más adelante, a partir de los Niveles Digitales de las distintas bandas de una imagen. Este análisis es de tipo puramente numérico. La conversión de esos Niveles Digitales en colores permite extraer los mismos resultados en base a la mera exploración visual, ya que presenta los valores de las bandas de un modo en que el ojo humano puede interpretarlos y analizarlos. Así, el análisis más inmediato que puede realizarse con una imagen es el de tipo visual, siempre que se tenga una representación adecuada de la imagen.

Las imágenes más sencillas contienen una única banda, y su representación en blanco y negro (escala de grises) es la forma habitual de mostrarlas. Las imágenes digitales habituales se dividen en tres bandas, denominadas *canales*, cada una de las cuales contiene la información sobre un color concreto. En el caso más frecuente del modelo RGB, ya conocido, estos colores son el rojo, el verde y el azul.

Las imágenes que tomamos con una cámara digital o una fotografía aérea en color escaneada y almacenada en un fichero, ambas utilizan este modelo de colores. Los periféricos de representación en los cuales las observamos, tales como una pantalla o un cañón de proyección, utilizan un esquema igual para reproducir un color dado. Este esquema se denomina de tipo *aditivo*, ya que forma el color sumando unas intensidades concretas de cada uno de los colores básicos. La figura 17.15 muestra un ejemplo de descomposición en canales RGB.

Según lo anterior, el color verde de, por ejemplo, una zona de vegetación, se descompone en una cierta intensidad de verde, una de rojo y una de azul, se almacena, y después se recrea en el monitor de un ordenador sumando dichas intensidades. Es decir, existe una relación directa entre el color *real* y el que vemos en pantalla.

A diferencia de esto, las imágenes de satélite multiespectrales no tienen el mismo significa visual, pero es interesante buscar de igual modo una forma de visualizarlas para poder llevar a cabo este primer análisis. Los Niveles Digitales de las distintas bandas no están relacionados con uno de los colores base de ningún modelo. Como sabemos, representan las intensidades correspondientes a una longitud de onda concreta, pero esta no tiene por qué ser la de un color básico o existir un modelo que se base en dichos colores. De hecho, ni siquiera tiene que corresponder a una longitud de onda dentro del espectro visible, ya que puede corresponder, por ejemplo, al infrarrojo. Asimismo, el número de bandas no ha de ser necesariamente de tres, y puede ser muy superior.

En estos casos, se toman tres bandas de entre las que componen la imagen, y se asocian estas a uno u otro de los canales del rojo, el verde y el azul. En función de las bandas y el esquema de asignación que se tome, la representación visual que se obtiene es muy distinta, y más o menos válida para uno u otro tipo de análisis.

De entre los posibles subconjuntos de tres bandas y el orden de asignación de estas, algunos presentan una mayor lógica de cara a la realización de los análisis más frecuentes. Así, resul-

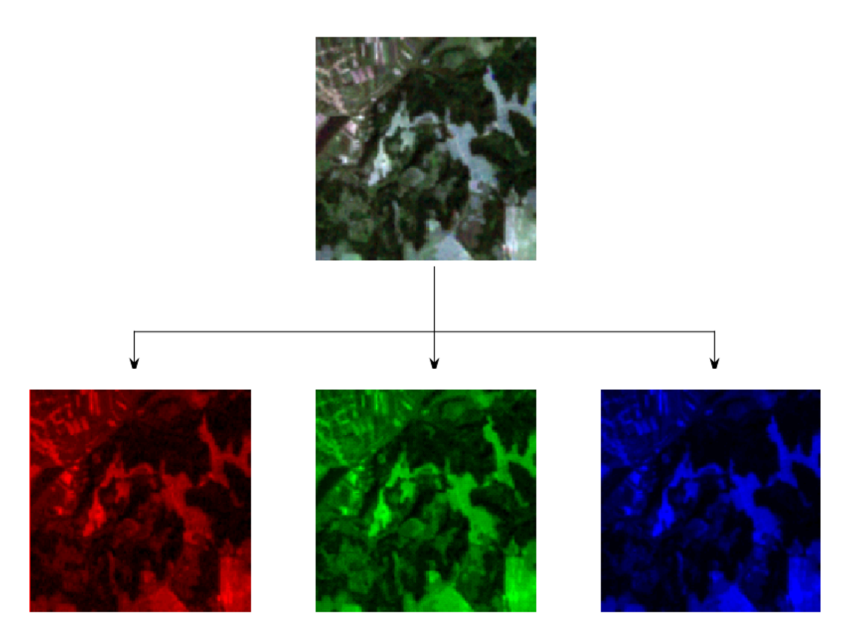


Figura 17.15: Descomposición de una imagen en color en sus componentes RGB.

ta fácil pensar que la opción más interesante es tratar de seleccionar aquellas tres bandas cuyas longitudes de onda sean más similares a las del rojo, el verde y el azul, de tal modo que la composición resultante sea lo más similar posible a la tonalidad real del terreno fotografiado. Para imágenes del sensor LANDSAT TM, por ejemplo, esto lleva a emplear las bandas 3 (rojo), 2 (verde) y 1 (azul).

En otros casos, no existen tales bandas, y puede o bien tomarse una banda alternativa, o proceder a efectuar combinaciones lineales de las bandas existentes. En este ultimo caso, se *simula* la banda que no está presente.

El cociente entre dos bandas dadas es también una solución habitual a utilizar para obtener los valores para los distintos canales.

Si se toman otras bandas distintas y se las asocia a un canal dado (pese a que realmente no contienen información acerca de la intensidad en la longitud de onda de dicho color), se obtiene lo que se conoce como composiciones en *falso color*.

Una composición en falso color habitual es la que utiliza la banda del infrarrojo cercano para el canal rojo, la del rojo para el canal azul y la del verde para el canal azul. El resultado es similar a de una fotografía infrarroja, y su interpretación en función de los distintos colores pude hacerse de la misma forma que en esta. Para el sensor LANDSAT TM, las bandas 4,2,1 o bien las bandas 3,2,1 son las que se deben usar para generar esta composición.

En general, pueden tomarse cualesquiera bandas para crear una composición en falso color. La elección de unas u otras y el canal al que se asocian debe estar en relación con la finalidad que se persiga, de forma que la composición resultante realce las características de mayor interés para la interpretación visual de la imagen. Como veremos más adelante en este capítulo, la actividad de la vegetación afecta notablemente a los valo-

res en el rango del infrarrojo, y son estos valores los que se utilizarán para los análisis relativos a dicha vegetación. Este mismo concepto puede aplicarse a la hora de la visualización, y si queremos discriminar la vegetación existente con un análisis puramente visual, será de interés incorporar la banda correspondiente a dicha longitud de onda al crear la composición en falso color, ya que aportará información a tal fin.

17.4.2. Operaciones morfológicas

Las operaciones morfológicas producen modificaciones de las *formas* presentes en la imagen, siendo estas formas las de los distintos *elementos* que encontramos en dicha imagen. En su concepción básica, estas operaciones se aplican sobre imágenes binarias (recordemos que estas pueden generarse en base a otro tipo de imágenes mediante las formulaciones vistas en la sección 17.3.1), las cuales, según vimos, contienen únicamente dos valores: uno de ellos que representa los citados elementos y otro que representa el *fondo* de la imagen.

Las operaciones morfológicas se aplican originalmente sobre imágenes binarias, pero existen igualmente versiones de los algoritmos correspondientes que pueden aplicarse sobre imágenes en escala de grises o incluso imágenes multibanda. A la hora de emplearlas dentro de un SIG, podemos aplicar este tipo de operaciones sobre cualquier capa ráster que contenga de algún modo *formas* definidas por conjuntos de píxeles. Tales capas pueden obtenerse como resultado de muchas de las operaciones que hemos visto en capítulos anteriores, y pueden también prepararse empleando algunos de los procedimientos presentados dentro de este mismo, tales como la detección de bordes. También, por supuesto, pueden ser imágenes más complejas que contengan este tipo de elementos, tales como las distintas zonas de cultivo de una zona dada, aunque estas requieren preparación o procesos previos. La *Clasificacion*, que veremos en la sección 21.2, es uno de tales procesos.

La importancia de estas funciones reside en que sirven para alterar los elementos de la imagen, siendo útiles para que otras operaciones posteriores se desarrollen de una forma más adecuada. Algunos de esos procesos posteriores que se benefician de la aplicación de operaciones morfológicas son la vectorización (que estudiaremos en el apartado 18.2) o la medición de longitudes en caso de elementos lineales.

Dos son las operaciones morfológicas básicas:

- Erosión
- Dilatación

Basadas en estas operaciones, existen otras derivadas, entre las cuales veremos tres:

- Apertura
- Cierre
- Esqueletización

Erosión y dilatación

Las dos principales operaciones morfológicas son la erosión y la dilatación. Pese a que sus nombres son ya sumamente descriptivos, veamos gráficamente en la figura 17.16 el efecto de ambas sobre una imagen base.

Como es fácil apreciar, la erosión reduce el tamaño de las formas, mientras que la dilatación tiene el efecto contrario, «engordándolas»[4].

La utilidad de estas operaciones es amplia, no solo por sí mismas, sino especialmente si consideramos que son la base de otras operaciones que resultan en transformaciones de gran interés

[4]Este efecto contrario en realidad hace que las operaciones sean una la *dual* de la otra. Esto quiere decir que aplicar una de ellas sobre los elementos de la imagen es equivalente a aplicar la otra sobre el fondo

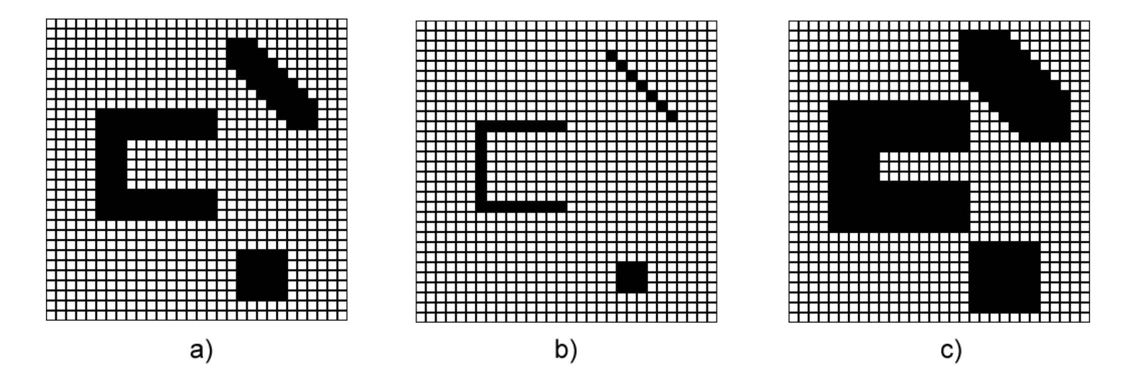


Figura 17.16: Resultado de aplicar operaciones de erosión (b) y dilatación (c) sobre una imagen base (a).

para ciertos cálculos muy habituales en el trabajo con un SIG.

En ambos casos, partimos de una imagen binaria y de un núcleo o *kernel*, que es una matriz de pequeño tamaño, del mismo modo que ya vimos en el caso de las convoluciones[5]. Las operaciones que aplicamos en este caso no son de tipo algebraico, a diferencia de las convoluciones, sino de tipo lógico, aunque desde nuestro punto de vista del álgebra de mapas, estos procesos son parte igualmente del grupo de operadores focales.

Como en muchas operaciones de tipo focal, la ventana más sencilla y habitual es la de tamaño 3×3. Por ejemplo, para el caso de la erosión, la siguiente es la ventana que utilizaremos.

1	1	1
1	1	1
1	1	1

Para efectuar dicha erosión desplazamos la ventana por todos los píxeles de los elementos que se quieren erosionar (que en una imagen binaria tendrán un valor 255 o 1 según hemos visto en el apartado dedicado a la segmentación de imágenes). Si con el anterior núcleo centrado en un píxel concreto todos los píxeles circundantes en la imagen coinciden con los valores de la ventana, entonces el píxel mantiene su valor 1 como píxel de un elemento. En caso contrario, pasa a tener un valor cero, pasando a ser un píxel del fondo.

El anterior núcleo aplicado según la metodología descrita hace que solo se mantengan con valor 1 aquellos píxeles que estén completamente rodeados de otros píxeles también de valor 1. Por ello, todos los píxeles del exterior de los distintos elementos quedan eliminados (pasan a formar parte del fondo), y al área de estos elementos disminuye.

El efecto de la erosión es mayor si se utiliza una ventana de mayor tamaño, tal como una 5×5, también con valores 1 en todas sus celdas. Este efecto más acusado puede reproducirse igualmente si se aplican sucesivos procesos de erosión con la ventana menor, reduciendo progresivamente el tamaño de los elementos.

Para el caso de la dilatación, podemos utilizar el mismo elemento estructural, pero el criterio es distinto. En el caso de la erosión se exigía una completa coincidencia entre la imagen y el ele-

[5]Estos núcleos se denominan también *elementos estructurales* o *sondas* dentro de la jerga relativa a este tipo de operaciones.

mento estructural. En este, únicamente se exige que al menos uno de los píxeles presente coincidencia. Esta operación, con el núcleo presentado, implica que todo píxel que se encuentre en contacto con algún píxel del objeto, pasara a formar parte de este. Con ello, los objetos crecen.

Apertura y cierre

Combinando las operaciones de erosión y dilatación, se definen otras nuevas. Las dos más básicas son la *apertura* y el *cierre*.

La apertura implica realizar una erosión seguida de una dilatación, ambas con el mismo elemento estructurante. Este proceso tiene como resultado un suavizado de los contornos, así como la eliminación de píxeles aislados.

Por su parte, el cierre también se componen de erosión y dilatación, pero esta vez en orden inverso. Es decir, primero se aplica la dilatación y posteriormente la erosión. Igual que antes, ambas operaciones utilizan el mismo elemento estructural.

La aplicación de una operación de cierre rellena los huecos y concavidades de los objetos que son menores en tamaño que el elemento estructural.

Una aplicación práctica de estas operaciones la podemos encontrar en la imagen 17.17. En ella la imagen inicial a) presenta una estructura lineal tal como una carretera, pero esta, a través de los distintos procesos aplicados para obtener la imagen binaria, ha «perdido» algunos píxeles, de forma que ya no existe una conexión correcta entre los distintos píxeles que la forman. La carretera ha quedado «rota» en fragmentes inconexos, que no representan realmente el objeto, y que van a dificultar el cálculo de parámetros acerca de este.

La operación de cierre, como hemos dicho, rellena de los huecos de la imagen. Este efecto es exactamente el que necesitamos en este punto, ya que incorporará a la carretera aquellos píxeles que faltan para establecer la conexión entre los distintos tramos y convertirla en un único objeto.

El resultado de aplicar una operación de cierre puede verse en la imagen b) dentro de la figura 17.17. Además de *cerrar* la vía, ha causado que el elemento en conjunto haya aumentado de tamaño y sea ahora más grueso. La forma de corregir este hecho (que no es adecuado para otras operaciones) es lo que veremos seguidamente con una nueva operación morfológica: la esqueletización.

Esqueletización

Hemos visto cómo resulta posible a partir de la aplicación de métodos sencillos reducir el tamaño de los distintos objetos de la imagen, y aplicando estos sucesivamente, reducir dicho tamaño más y más hasta el punto deseado. Utilizando ideas similares podemos reducir un objeto de una imagen binaria hasta un conjunto mínimo de píxeles que preserven la forma y las propiedades topológicas del objeto original, habiendo eliminado el mayor numero posible de ellos, que quedarían como píxeles de fondo. En particular, un conjunto de un único píxel de ancho es la estructura mínima que puede obtenerse.

Este conjunto mínimo de píxeles delimitan el «eje» del objeto, y se conocen como el *esqueleto* de este. La obtención de este esqueleto es de gran utilidad, pues al constituir un objeto de máxima simplicidad, facilita buena parte de las operaciones que se pueden realizar posteriormente.

Para ver un caso práctico, volvamos a la figura 17.17. La imagen b) que hemos obtenido tras el proceso de cierre representa, según convenimos, una carretera, pero el tamaño de celda de la misma es superior al ancho de la carretera. Por esta razón, el objeto que define la carretera tiene un ancho de varias celdas.

Un parámetro sencillo tal como la longitud de la vía no resulta difícil de medir manualmente



Figura 17.17: a) Imagen original con un objeto lineal desconectado. b) Objeto conectado tras la aplicación de una operación de cierre. c) Esqueleto del anterior objeto obtenido mediante la aplicación de operaciones morfológicas.

en esa imagen b), pero automatizar ese proceso sobre la misma imagen plantea ciertas complicaciones. Puesto que la carretera también tiene un cierto ancho, la longitud que queremos calcular no queda definida unívocamente, al menos no de una forma inmediata. Baste ver que, sin salirnos de los píxeles negros (los de carretera), podemos ir de un extremo a otro de esta por muchas rutas distintas, saltando de píxel en píxel y obteniendo en estos recorridos longitudes distintas.

El problema estriba en que, aunque tratemos conceptualmente a esa carretera como un elemento lineal —es decir, sin ancho—, en realidad sí que tiene un ancho dado, a pesar de que este es sensiblemente menor que su longitud. Si observamos ahora la imagen c) dentro de la anterior figura, veremos que estos problemas ya no existen. Solo hay una forma de recorrer el conjunto de píxeles de un extremo a otro. La medición puede claramente automatizarse, y el algoritmo encargado de ello será mucho más simple.

El esqueleto de la carretera mantiene así las propiedades que nos interesan, tales como la longitud, pero de una manera más adecuada para el análisis. De hecho, mantiene incluso la topología, la cual habíamos «corregido» aplicando la operación de cierre.

El cálculo del esqueleto puede llevarse a cabo aplicando operaciones similares a las que ya conocemos. Estas operaciones se denominan de *adelgazamiento*[6], pues su efecto es, en efecto, el de adelgazar el objeto. En esencia, puede entenderse como un proceso de erosión aplicado repetidamente hasta que al final quede un objeto de un único píxel de ancho, y garantizando que no se pierda la topología original (no surjan desconexiones). Un algoritmo habitual de adelgazamiento es el propuesto por [237].

Otra forma de entender el cálculo del esqueleto de un objeto es a partir de distancias. A cada píxel de dentro del objeto se le puede asignar el valor de la distancia al píxel de fondo más cercano. Los puntos que forman parte del esqueleto cumplen la condición de ser máximos locales de distancia. Es decir, que todos los píxeles circundantes están más cerca del borde que el píxel en cuestión[7] .

En ambos casos, la obtención de un esqueleto, aunque pueda ser diferente según la formulación aplicada, mantiene las propiedades del objeto ori-

[6]Habitualmente citadas en su denominación en inglés: *thinning*

[7]Aunque sean formulaciones que hayan surgido en campos distintos, este uso de distancias es similar en su forma a lo que veremos en el capítulo 20 acerca del cálculo de costes. Los puntos de destino que entonces estudiaremos son en este caso los píxeles del fondo

ginal, pero reduciendo este a una expresión mínima y más sencilla de estudiar.

17.4.3. Estimación de parámetros físicos

Los Niveles Digitales de las distintas bandas de una imagen expresan la reflectividad para una longitud de onda dada. Es decir, tienen un significado físico. El hecho de que a partir de una fuente de iluminación única se presenten en la capa distintos Niveles Digitales es debido a que los distintos elementos de un área de estudio reflejan la luz de dicha fuente de forma diferentes.

Una zona boscosa no refleja la luz igual que un embalse, y dos zonas boscosas tampoco lo harán igual si sus árboles son de distintas especies o tienen un estado fitosanitario diferente. Si los elementos de una área de estudio condicionan la radiación reflejada en las distintas longitudes de onda, podemos utilizar esta radiación reflejada (la que se recoge en los Niveles Digitales de esas bandas) para deducir las propiedades de los elementos.

Atendiendo a lo anterior surge el concepto de *firma espectral*. La firma espectral es el conjunto de reflectancias en las distintas longitudes de onda que presenta un objeto o material dado. Dicho de otro modo, la firma espectral es la «huella dactilar» que caracteriza a cada objeto, y que deriva directamente de sus propiedades y de cómo estas influyen en la forma en la que dicho objeto refleja la radiación incidente.

En la figura 17.18 pueden verse algunas firmas espectrales características, teniéndose así una idea de las diferencias que pueden existir, y también de cómo estas pueden utilizarse para el análisis de imágenes.

Más detalladamente, y para un píxel caracterizado por su firma espectral, dos son los tipos de propiedades que se pueden extraer:

- ¿Qué hay en ese píxel? Puesto que los distintos objetos van a presentar firmas espectrales particulares, pueden caracterizarse las firmas más habituales de una serie de objetos tipo, y después tratar de identificar a cuál de ellas se asemeja más la presente en un píxel concreto.
  Los fundamentos teóricos de este proceso, que son de tipo genérico y pueden aplicarse a otras fuentes de datos además de las imágenes, se tratarán en el capitulo 21. En este capítulo principalmente veremos los aspectos relacionados con la identificación de factores físicos a partir de imágenes, que podrán emplearse en conjunción con los que se verán entonces.
  En general, la clasificación de elementos del medio hace uso de las denominadas *bases de datos espaciales*[8], que recogen las características de dichos elementos expresadas en función de las energías recibidas en las distintas longitudes de onda.
  Más información sobre bases de datos de firmas espectrales puede encontrarse en [238].
- ¿Cómo es lo que hay en ese píxel? Conociendo qué tipo de elemento encontramos en un píxel (por ejemplo, una superficie de agua como un embalse), pueden estimarse asimismo las cualidades particulares del mismo. Características como la temperatura de ese agua tienen efecto sobre la firma espectral, y pueden deducirse mediante formulaciones desarrolladas a tal efecto.

Según sea la propiedad que queremos conocer o el tipo de elemento que pretendemos detectar en el terreno, será una u otra longitud de onda (es decir, uno u otro sensor y una u otra banda de las imágenes generadas por este) la que aporte

[8]No ha de confundirse esto con la tecnología de bases de datos con capacidad de manejo de elementos espaciales, la cual ya hemos visto en otros capítulos. En este caso no nos referimos a la tecnología, sino a un conjunto particular de datos que almacenan esas firmas espectrales de una serie de elementos conocidos.

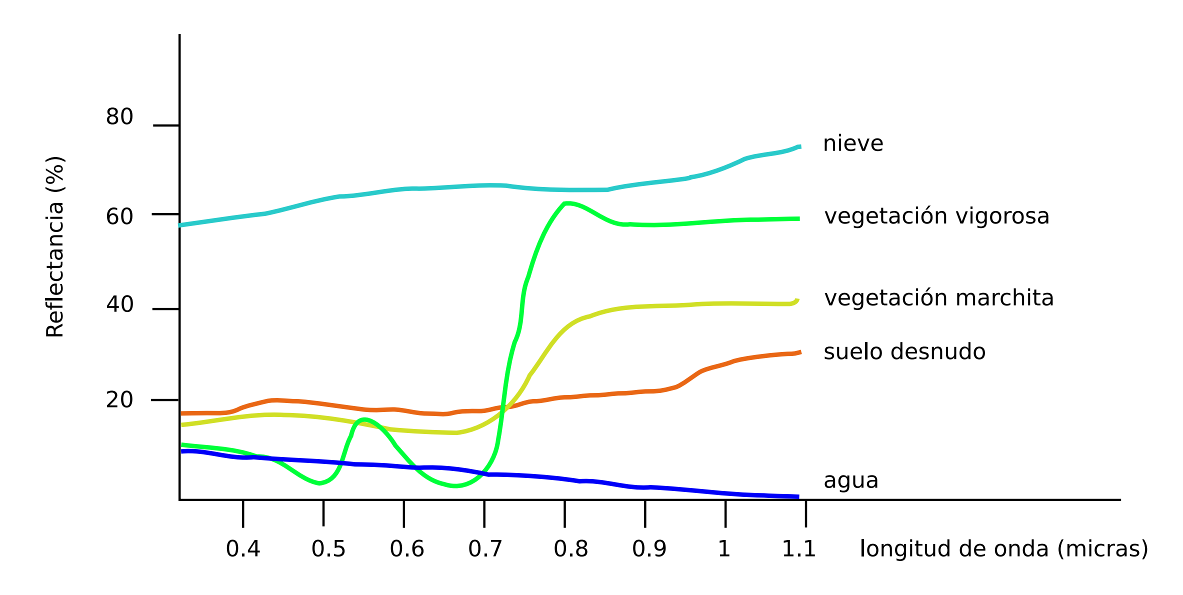


Figura 17.18: Firmas espectrales tipo de algunos elementos habituales

una información más relevante. En el cuadro 17.1 se muestran de forma resumida las aplicaciones principales de las distintas longitudes de onda, en función de los procesos en los que dichas radiaciones toman parte.

Veremos a continuación con algo más de detalle cómo deducir propiedades de los dos tipos anteriores en relación con tres de los elementos básicos del estudio del medio: el suelo, la vegetación y el agua.

Parámetros de la vegetación

La vegetación es uno de los elementos que mayor atención han recibido en el ámbito del análisis de imágenes. Por ello, existen muchos parámetros que pueden emplearse para obtener información sobre la vegetación a partir de imágenes.

Uno de los parámetros principales son los denominados *índices de vegetación*. Estos índices permiten detectar la presencia de vegetación en una zona y la actividad de esta, ya que sus valores están en relación con dicha actividad. Para ello, utilizan los valores de reflectancia correspondientes a las distintas longitudes de onda, interpretando estas en relación con la actividad fotosintética.

En términos generales, un índice de vegetación puede definirse como un parámetro calculado a partir de los valores de la reflectividad a distintas longitudes de onda, que pretenden extraer de ellos la información relativa a la vegetación, minimizando las perturbaciones debidas a factores tales como el suelo o la atmósfera [239].

El conocimiento del proceso de fotosíntesis ayuda a comprender la base de los índices de vegetación. Por regla general, las plantas absorben radiación en las regiones visibles del espectro, en particular en la región del rojo, siendo esta radiación la que emplean para su obtención de energía. Por el contrario, las radiaciones de menor frecuencia tales como las del infrarrojo no son útiles para el desarrollo de la actividad fotosintética, y la absorción de estas provocaría tan solo un aumento de la temperatura de la planta que podría ser perjudicial para la misma. Por ello,

Región del espectro	Procesos	Aplicaciones
Rayos X	Procesos atómicos	Detección de elementos radiactivos
Ultravioleta	Procesos electrónicos	Presencia de H y He en la atmósfera
Visible e IR cercano	Vibración molecular	Composición química de la superficie. Propiedades biológicas
IR medio	Vibración y rotación molecular	Composición química de la superficie y la atmósfera
IR térmico	Emisión térmica	Temperatura de la superficie y la atmósfera
Microondas	Rotación molecular y emisión térmica	Composición química de la atmósfera. Propiedades físicas de la superficie.

Cuadro 17.1: Esquema de regiones del espectro, procesos con los que interactúan y aplicaciones habituales en función de estos.

la mayor parte de estas frecuencias se reflejan en la planta.

Como resultado de lo anterior, y en caso de que exista vegetación, las bandas de imágenes correspondientes a la zona del infrarrojo van a presentar altos valores de reflectancia, mientras que los de la banda del rojo serán bajos. Esto permite localizar las zonas de vegetación mediante la utilización de dichas imágenes. Este es el fundamento de los índices de vegetación.

Sin necesidad de definir aún ningún índice, podemos ver claramente este hecho en la figura 17.19. En ella, se ha creado una imagen en falso color empleando las bandas del infrarrojo cercano, el rojo y el verde para los canales del rojo, verde y azul respectivamente (los del modelo RGB)[9]. Debido a lo anterior, las zonas con cultivos con alto vigor vegetativo aparecen con un tono rojo muy marcado, que permite rápidamente identificarlas.

Dependiendo del tipo de sensor del cual provengan las imágenes a analizar, serán unas u otras bandas las que deban emplearse como bandas del rojo e infrarrojo cercano. En el caso particular del sensor LANDSAT, por ejemplo, la banda del rojo es la 3, mientras que la del infrarrojo cercano es la 4.

Según cómo se formulen los índices en base a los valores de las anteriores bandas, podemos distinguir dos grupos:

- Formulaciones empleando únicamente las bandas del rojo y el infrarrojo cercano. Estos se conocen como *índices basados en distancia*, o *intrínsecos*.
- Formulaciones empleando además los parámetros de la denominada *línea del suelo*. Estos índices se conocen como *índices basados en pendiente*.

Veremos más adelante qué representa esa *línea de suelo* y cómo calcularla. Por ahora, veamos algunos de los principales índices del primer grupo.

El más importante de los índices de vegetación es el *Índice de Vegetación de Diferencia Nor-*

[9]Este esquema es de uso habitual y se denota como RGB = NRG, siendo N = Near infrared (infrarrojo cercano), R = Red (rojo) y G = Green (verde). Como regla mnemotécnica, la lectura en ingles del acrónimo NRG suena parecida a la del vocablo inglés *energy* (energía), y es también común referirse así a él.

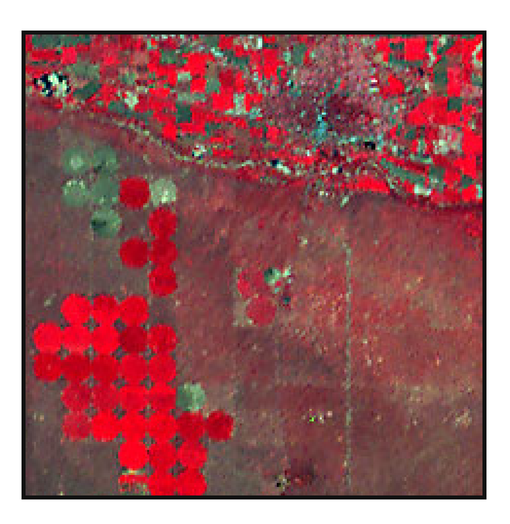


Figura 17.19: La creación de una imagen de falso color con un esquema RGB=NRG permite localizar las zonas de vegetación por su coloración roja intensa. Esto es debido a la alta reflectancia en las banda del infrarrojo visible y la baja reflectancia en el rango visible, causadas ambas por la actividad fotosintética.

malizada (NDVI, *Normalized Differential Vegetation Index*) [240], cuya expresión es la siguiente:

$$\mathrm{NDVI} = \frac{\mathrm{IR} - \mathrm{R}}{\mathrm{IR} + \mathrm{R}} \qquad (17.4.1)$$

siendo R e IR las reflectancias correspondientes al rojo e infrarrojo respectivamente.

Los valores de este índice se encuentran dentro del intervalo (-1,1). Valores altos de NDVI indican la presencia de vegetación.

Además de permitir localizar las zonas con presencia de vegetación, el NDVI puede emplearse para conocer otros parámetros relativos a la vegetación, ya que estos también condicionan la forma en que esta refleja la radiación. Factores tales como el vigor vegetativo, el estado fitosanitario o el contenido en agua de las hojas influyenen los procesos fotosintéticos. Esto tiene un efecto directo sobre las radiaciones reflejadas para las distintas longitudes de onda, y muy particularmente en las empleadas para la definición de los índices de vegetación. Por ello, son una valiosa fuente de información acerca de las variables propias de la vegetación.

Otros parámetros como el *Índice de Área Foliar*[10] o el factor C de la USLE pueden calcularse en primera aproximación a partir del NDVI, siendo abundante la literatura en este sentido. Por ejemplo, para el factor C, la siguiente expresión genérica puede servir para estimarlo en función exclusivamente del NDVI [241],

$$C = e^{-\alpha \frac{\mathrm{NDVI}}{\beta - \mathrm{NDVI}}} \qquad (17.4.2)$$

siendo valores habituales $\alpha = 2$ y $\beta = 1$.

En la figura 17.20 puede verse el aspecto de una capa de NDVI. Si se compara con la composición de color mostrada anteriormente puede observarse cómo las zonas sin vegetación presentan tonos oscuros (valores bajos de NDVI) y las de vegetación tonos claros (valores altos de NDVI).

El número de índices de vegetación dentro de este primer grupo es muy elevado, todos ellos con una formulación similar a la del NDVI y tratando

[10] Leaf Area Index(LAI)

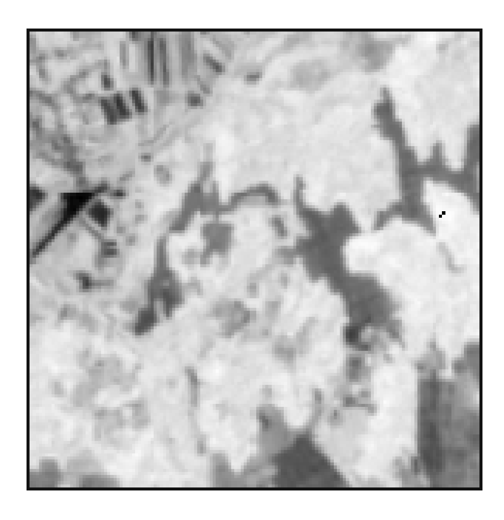


Figura 17.20: (NDVI)

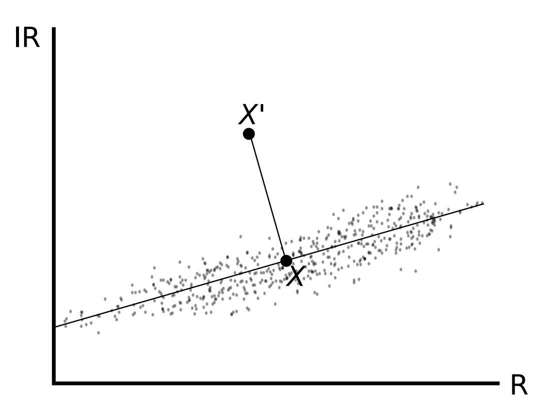


Figura 17.21: Representación gráfica de la línea de suelo, un punto correspondiente a suelo desnudo (X) y un punto correspondiente a la presencia de cierta vegetación (X')

de mejorar ciertos aspectos de este. Algunos de los más habituales son los mostrados a continuación.

- Índice de Vegetación de Razón Normalizada (Normalized Ratio Vegetation Index)

$$\mathrm{NRVI} = \frac{\frac{\mathrm{R}}{\mathrm{IR}} - 1}{\mathrm{IR} + \mathrm{R} + 1} \tag{17.4.3}$$

- Índice de Vegetación Transformado (Transformed Vegetation Index)

$$\mathrm{TVI} = \sqrt{\mathrm{NDVI} - 0{,}5} \quad ; \quad \mathrm{NDVI} > 0{,}5 \tag{17.4.4}$$

- Índice de Vegetación Transformado de Thiam (Thiam's Transformed Vegetation Index)

$$\mathrm{TTVI} = \sqrt{|\mathrm{NDVI} + 0{,}5|} \tag{17.4.5}$$

Como hemos dicho, las bandas del rojo y el infrarrojo cercano se corresponden en el caso del sensor LANDSAT TM con las bandas 3 y 4 respectivamente. La formulación original del NDVI se basaba en datos de dicho sensor, por lo que es habitual encontrar este bajo la denominación NDVI43. Si se emplea una notación similar, pueden definirse nuevos índices que mantienen la expresión del NDVI, pero utilizan distintas bandas como base de su cálculo. Algunos de estos índices derivados han sido estudiados y documentados en la literatura, y es posible asignarles una interpretación del mismo modo que con el modelo base del NDVI. Así, [242] define el NDVI54 y lo relaciona con el estrés hídrico de la vegetación y [243] define el NDVI75, el cual presenta una correlación apreciable con la biomasa existente.

Para el cálculo de los índices de vegetación del segundo grupo se requiere, como ya hemos visto, el cálculo de una recta denominada *línea de suelo*. Esta recta viene definida por su pendiente b y el punto a en el que corta al eje de ordenadas, siendo estos valores los que se incorporan a las formulaciones de los índices de este grupo.

El fundamento en el que reside el uso de la línea de suelo y su incorporación a los índices de vegetación puede comprenderse con facilidad si se observa la figura 17.21. Los puntos de la gráfica representada son los correspondientes a los valores en las bandas del rojo y el infrarrojo para los píxeles sobre zonas de suelo desnudo. A

estos se ajusta la recta r mostrada, y los puntos sobre la misma tales como el punto X se asocian al suelo desnudo.

Para ese mismo suelo, una cierta cobertura vegetal haría que el valor correspondiente tuviera una menor reflectancia en el rojo y mayor en el infrarrojo cercano, alejándose de la línea de suelo perpendicularmente. Cuanto mayor es la cobertura vegetal y mayor la actividad fotosintética, mayor será esta distancia entre el punto y la línea de suelo, y por lo tanto su medida puede constituir un índice de vegetación válido. Esta es la razón por la que estos índices, como ya se dijo, se conozcan como índices basados en distancia.

El cálculo de la línea de suelo se lleva a cabo siguiendo los pasos enunciados a continuación:

- Se calcula un índice de vegetación basado en pendiente tal como el NDVI.
- Se identifican las zonas sin vegetación (suelo desnudo) en base al anterior. Estas tendrán valores negativos próximos a -1.
- Se calcula una regresión lineal entre los valores del rojo y del infrarrojo, solo con los píxeles de las anteriormente definidas zonas sin vegetación. Se obtendrá una expresión de la forma $y = a + bx$. Según sea el índice de vegetación, deberá tomarse como variable independiente o bien la reflectancia correspondiente al rojo, o bien la del infrarrojo.

Una vez realizado el proceso anterior, con los coeficientes a y b pueden formularse nuevos índices de vegetación para estimar la distancia anteriormente descrita. El índice de vegetación original dentro de este tipo, del cual derivan los restante, es el conocido como *Índice de Vegetación Perpendicular*(PVI, *Perpendicular Vegetation Index*) [244].

Este índice, no obstante, es complejo y además no aporta información sobre si la distancia se da hacia el otro lado de la línea de suelo (agua en lugar de vegetación), por lo que se han propuesto otros más sencillos y eficaces. Entre ellos, cabe destacar los siguientes:

- [245]

$$\mathrm{PVI} = \frac{b \cdot \mathrm{IR} - \mathrm{R} + a}{\sqrt{b^2 + 1}} \tag{17.4.6}$$

- [246]

$$\mathrm{PVI} = \frac{(\mathrm{IR} - a)(\mathrm{R} + b)}{\sqrt{a^2 + 1}} \tag{17.4.7}$$

.

- [247]

$$\mathrm{PVI} = a \cdot \mathrm{IR} - b \cdot \mathrm{R} \tag{17.4.8}$$

Un último índice de formulación similar al NDVI es el SAVI (*Soil Adjusted Vegetation Index*) [248]. Su expresión es la siguiente:

$$\mathrm{SAVI} = \frac{\mathrm{IR} - \mathrm{R}}{\mathrm{IR} + \mathrm{R} + L}(1 + L) \tag{17.4.9}$$

La incorporación del parámetro L minimiza las perturbaciones causadas por el suelo y debe estimarse en función de este. El autor recomienda un valor de 1 para zonas de vegetación escasa, 0.5 en zonas de densidad media, y 0.25 si la densidad de la vegetación es elevada.

Parámetros del suelo

El suelo es otro elemento del cual pueden saberse mucho a partir de imágenes de satélite. Junto con la inspección visual, que nos permite analizar la capa superficial del suelo, las reflectancias a distintas longitudes de onda permiten profundizar en este y estudiar también la estructura y propiedades de los horizontes inferiores.

Si en el caso de la vegetación es la actividad fotosintética la que condiciona la reflectancia y permite estudiar la presencia y propiedades de la vegetación, en el caso del suelo son cuatro los

factores principales que influyen: la composición mineralógica del suelo, la humedad, la textura y la materia orgánica.

Resulta claro que cada mineral va a presentar una comportamiento distinto frente a la radiación, y ello puede utilizarse para inferir la composición del suelo. Puede realizarse una clasificación de los suelo partir de imágenes de satélite, para lo cual existen bases de datos con las firmas espectrales de los principales minerales.

Respecto a la humedad, se presentan bandas de absorción marcadas alrededor de los 1,4 y 1,9 μm. Por su parte, las bandas del rojo y el infrarrojo cercano son indiferentes a la cantidad de humedad presente en suelo. Estas bandas, que utilizábamos para la definición de un tipo particular de índices de vegetación, también sirven para caracterizar el suelo. La regresión entre ellas da lugar a la ya conocida *línea del suelo*, cuyos valores sirven para definir dicho suelo.

La materia orgánica, por su parte, influye sobre la capacidad de retención de humedad del suelo, por lo que indirectamente afecta a las propiedades ópticas de este y condiciona los valores de reflectividad.

Por último, la textura influye sobre la reflectancia al modificar la forma en que la luz incide y se refleja sobre el suelo. La rugosidad y tamaño de las partículas causan que estos procesos se den de un modo u otro, y afectan así a las imágenes.

Junto con los parámetros del suelo que pueden estudiarse directamente a partir de la radiación reflejada por este, otros pueden analizarse empleando elementos adicionales como, por ejemplo, la vegetación, de la cual ya hemos hablado con anterioridad. Aunque se requiere un conocimiento exhaustivo de las condiciones existentes, y debe complementarse su análisis mediante imágenes con el estudio de campo, las propiedades de las plantas tales como verdor o coloración pueden relacionarse con las propiedades del suelo, esencialmente la disposición de nutrientes en este.

Parámetros del agua

El estudio de los elementos hidrográficos (cauces, lagos, etc.) puede llevarse a cabo a partir de imágenes desde distintos puntos de vista.

El uso más inmediato es la delimitación de dichos elementos, ya que resulta sencillo identificar estos y sus límites. Las bandas del visible y el infrarrojo cercano son las más adecuadas para esta tarea, pues en ellas las superficies de agua se presentan de forma más conspicua.

Junto con la localización de los elementos hidrográficos, las propiedades del agua en los mismos son de igual modo materia de estudio. Así, la región espectral del infrarrojo térmico aporta información sobre la temperatura del agua, mientras que la región del ultravioleta permite la localización de compuestos químicos tales como aceites u otros contaminantes.

Puesto que, como hemos visto, la presencia de vegetación influye sobre la reflectancia, podemos servirnos de algunos conceptos como los estudiados en puntos anteriores para estudiar el contenido en algas y otros organismos fotosintéticos de una masa de agua. Otros elementos en suspensión, tanto de naturaleza orgánica como inorgánica, influyen de igual modo sobre la reflectancia, y permiten que la turbidez del agua pueda ser monitorizada en base a imágenes satelitales.

Mención aparte merece el estudio de mares y océanos, existiendo satélites como eñ NOAA especialmente desarrollados para este propósito. El estudio de las temperaturas permite controlar elementos como las corrientes marinas, entre otros. La superficie oceánica es también objeto de análisis, pudiendo extraerse valores de rugosidad y altura de las olas. Indirectamente, otros parámetros como la velocidad del viento, no relativos al agua, pueden derivarse a partir de los anteriores.

Por último, mencionar que el estudio del agua en estado solido (nieve, hielo) comparte algunos de los conceptos anteriores, y puede, por ejemplo, monitorizarse el estado del manto de nieve en base a los valores de reflectancia en la región espectral del infrarrojo medio. Es de reseñar que en muchos casos la presencia de nieve satura los sensores debido a los altos valores de reflectancia que presenta en ciertas bandas.

Transformacion *Tasseled Cap*

La transformacion Tasseled Cap [249] guarda gran similitud operacional con el *análisis de componentes principales*, el cual veremos en el apartado 21.5, siendo un caso particular de este. Se trata de obtener una nueva serie de bandas a partir de las del conjunto original como combinación lineal de estas, cada una de ellas con un significado bien definido. Este significado guarda relación principalmente con la vegetación, aunque se trata aquí por separado y no en la sección anterior correspondiente, por no ser una formulación únicamente dirigida a este tipo de análisis.

Los coeficientes de las combinaciones lineales que expresan la transformación se encuentran tabulados para los sensores más habituales tales como LANDSAT TM+ o MSS, y en función de estos también puede variar el número y significado de las bandas resultantes.

El caso más frecuente es la utilización de bandas del sensor LANDSAT TM, empleándose todas ellas excepto la del infrarrojo térmico, es decir, un total de seis bandas. Las bandas resultantes (3) corresponden a los ejes de la línea de suelo (que expresa la humedad), la vegetación (el verdor) —recuérdese que este es perpendicular al anterior— y el brillo. Los valores de estas bandas por separado tienen ya una interpretación relacionada de forma directa con las variables físicas del medio.

17.4.4. Detección de elementos

Uno de los procesos más interesantes basados en imágenes es la localización y definición de los distintos elementos de una zona de estudio. Estos pueden ser elementos zonales tales como lagos o áreas con una vegetación concreta, o bien otros como carreteras o caminos, líneas de alta tensión o árboles puntuales.

Como ya sabemos, las imágenes nos permiten una actualización mucho más sencilla de las bases de datos espaciales, pero ello requiere extraer de ellas la información necesaria para actualizar dichas bases de datos. Cuando el proceso de extracción de información se automatiza, la actualización de las bases de datos lo hace de igual modo (al menos en cierta medida), aprovechándose realmente todo el potencial que las imágenes contienen para mantener actualizados los datos geográficos de forma sencilla.

La clasificación de imágenes, la cual veremos en el apartado 21.2, es una herramienta válida a estos efectos, ya que permite asignar una categoría a cada uno de los píxeles de una imagen, permitiéndonos convertir una imagen en una capa de tipos de suelo o usos de suelo, entre otras, y provee una solución automatizada para ello. Por su gran importancia, la veremos con gran detalle en un capítulo aparte.

Este tipo de procesos, sin embargo, no son siempre suficientes para la creación de nuevas capas. A veces se necesita no solo clasificar, sino también *detectar* los distintos elementos que componen la zona que estudiamos, en base principalmente a su forma o aspecto. Como veremos, la clasificación utiliza las propiedades de cada píxel para asignarle una categoría, pero no tanto las de forma.

Decíamos al inicio del capítulo que podemos, a partir de imágenes, identificar elementos y su características. Estos, sin duda, están relacionados, y las características nos pueden servir pa-




Figura 17.22: Identificación automatizada de edificios a partir de una imagen aérea (cortesía de Visual Learning Systems, Inc)

ra la propia identificación del elemento. Esta es la forma de operar del proceso de clasificación, pero las características no son suficientes. Por ejemplo, para buscar carreteras en una imagen podemos buscar zonas asfaltadas (saber si la zona es de asfalto lo podemos hacer a partir de las características de los píxeles, del mismo modo que visualmente lo identificaríamos por su color), pero pueden existir explanadas asfaltadas que no constituyan una carretera. Lo que nos lleva a determinar que una zona constituye una carretera es, en conjunto con su características de color, la forma lineal que esta tiene, con lo cual el análisis de las cualidades del píxel no resulta suficiente.

Del mismo modo, extraer la localización de cada árbol dentro de una masa arbórea también requiere identificar estos por la forma de sus copas en la imagen, al igual que, por ejemplo, la localización de edificios (Figura 17.22).

En cierta medida, el proceso de clasificación se basa de forma global en una concepción ráster de los resultados a obtener, mientras que la extracción de elementos enfoca dichos resultados como entidades vectoriales (que son, precisamente, esos elementos a detectar). Por ello, tiene en realidad mucha relación con todo lo que veremos en el capítulo 18, relativo a la creación de capas vectoriales. En conjunto con los conceptos que veremos entonces, aplicamos también los que conocemos de este capítulo —tales como las firmas espectrales—, así como las ya mencionadas ideas sobre clasificación que igualmente se detallarán en breve.

Puesto que la forma es el componente base para la detección de elementos, existe una relación lógica con aquellas operaciones que denominábamos *morfológicas*, ya que estas operan sobre imágenes y modifican la forma con que dichos elementos aparecen en ellas. Viendo el efecto de este tipo de operaciones, entenderemos fácilmente que pueden servirnos de ayuda para una detección de elementos más precisa, pudiendo emplearse en conjunto con otras técnicas para la preparación de las imágenes antes del propio proceso de detección. O bien, sencillamente, para la aplicación de los algoritmos que encontraremos en el capítulo 18.

A día de hoy, la detección de elementos no presenta el mismo nivel de automatización que otras tareas dentro de los SIG, y no es del to-

do completa. El exhaustivo conocimiento que se requiere de los elementos a detectar o la presencia de ruido en las imágenes son algunos de los factores que dificultan obtener resultados satisfactorios si se automatiza todo el proceso. Aun así, los avances en este sentido son de sumo interés por el gran ahorro de tiempo y costes que suponen para el mantenimiento de bases de datos espaciales actualizadas.

Las técnicas que se emplean en este campo son muy avanzadas e incluyen generalmente elementos de inteligencia artificial que permiten al sistema aprender e ir mejorando paulatinamente los resultados. Su presencia en los SIG actuales no se encuentra muy extendida, pero representan herramientas muy valiosas para sacar todo el partido posible a la gran cantidad de imágenes aéreas y de satélite de las que disponemos hoy en día.

17.5. Resumen

Las imágenes son uno de los elementos base de los Sistemas de Información Geográfica, y su análisis y tratamiento es una pieza clave dentro de los procesos que pueden llevarse a cabo en estos. Dicho análisis incluye ideas propias del tratamiento de imágenes digitales, en las que no se considera la naturaleza particular de las imágenes que se usan en un SIG, así como otras en las que sí se considera de forma directa la componente espacial y el significado físico de los valores que se recogen en cada píxel.

El tratamiento de las imágenes debe comenzar con la preparación y corrección de estas, eliminando los efectos no deseados que los aparatos de registro o la propia atmósfera, entre otros factores, hayan podido introducir en dichas imágenes.

La imagen corregida puede ser mejorada mediante algoritmos que hagan más patente la información que contiene, realzando esta tanto de cara a su mero análisis visual como para su posterior análisis mediante otras formulaciones especificas. Estas otras formulaciones permiten obtener, entre otros resultados, caracterizaciones de elementos del medio como el suelo, la vegetación o el agua.

Creación de capas vectoriales

18.1. Introducción

La conversión de datos espaciales desde el modelo vectorial al modelo ráster nos servía para obtener un dato espacial mucho más adecuado para el análisis. Para variables continuas tales como la elevación, los datos vectoriales resultan más difíciles de analizar, por lo que el tipo de conversión que vimos era un paso previo para poder acceder a toda la potencia de análisis de un SIG.

Las ventajas del modelo ráster, que fueron comentadas con detalle en el capítulo 5, lo convierten en la opción preferida a la hora del análisis de variables continuas, pero esta no es, ni mucho menos, la única tarea que podemos realizar en un SIG. Para muchas de esas restantes tareas, el modelo vectorial es preferible, pues, como ya vimos, también tiene su particular serie de ventajas.

Si en el capítulo 14 estudiábamos como convertir los datos vectoriales en datos ráster, debemos ver de igual forma cómo proceder en sentido inverso cuando esto sea posible, y mostrar qué formas vectoriales del dato espacial resulta posible obtener a partir de una capa ráster. El proceso que vamos a ver no es sencillamente el inverso de lo entonces presentado, sino que implica otra serie de operaciones.

Así, la conversión de puntos en una capa ráster, que nos ocupó la mayoría de paginas en aquel capítulo, no tiene un equivalente en la conversión de una capa ráster en una de puntos, pues esta segunda operación resulta obvia con los conocimientos que ya tenemos. Basta aplicar las ecuaciones vistas en 15.3 para calcular los valores de la variable en dichos puntos.

Otras operaciones, sin embargo, tales como convertir una capa ráster en una capa vectorial no de puntos sino de isolíneas, sí que requieren algo más de explicación. Más aún, resulta importante mostrar al lector que esos procesos existen y son parte de los procesos que un SIG nos ofrece, para que pueda incorporarlas a su batería de herramientas.

La coexistencia de los datos vectoriales y ráster no solo es una realidad, sino que en muchas ocasiones es una necesidad. A lo largo de las distintas fases de trabajo dentro de un proyecto SIG, un mismo dato puede emplearse de varias formas distintas con objeto de satisfacer las distintas necesidades que aparezcan. Como muestra la figura 18.1, para una capa de elevaciones, las conversiones entre modelos de datos aparecen en diversos puntos dentro de un ciclo de trabajo habitual. Una de ellas es una conversión de vectorial a ráster tal y como las que vimos en el capítulo anterior. Las restantes son en sentido inverso (de ráster a vectorial), como las que trataremos en el presente.

En el esquema de la figura, partimos de un mapa escaneado, el cual es una imagen donde se pueden ver las curvas de nivel con sus valores de

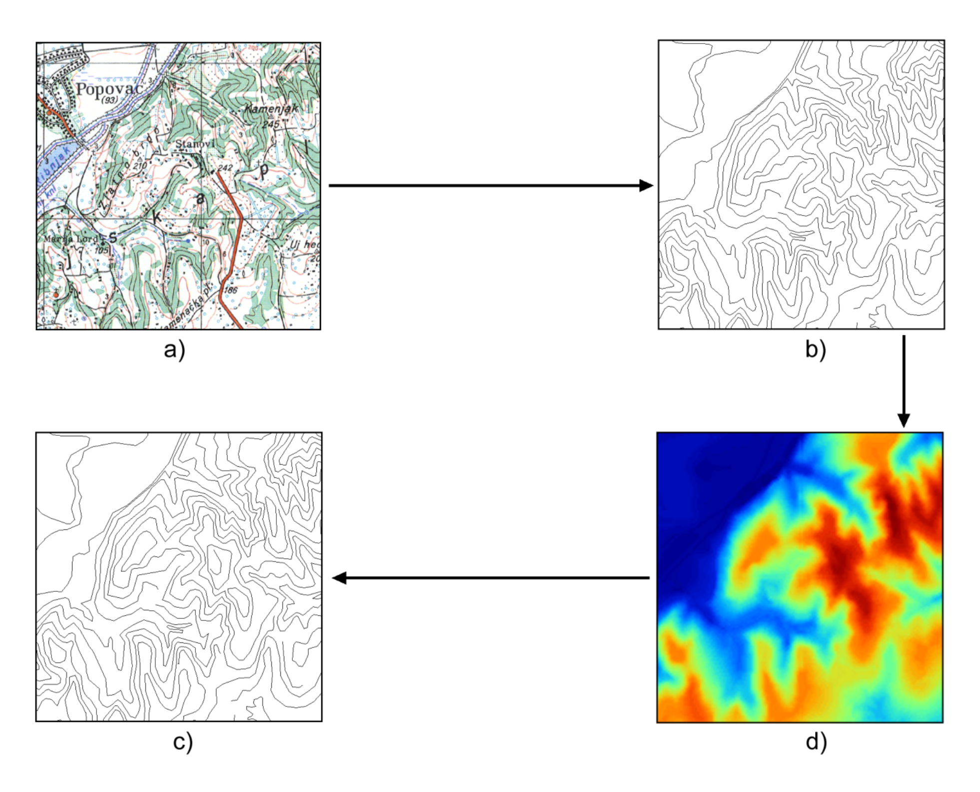


Figura 18.1: Distintas conversiones de modelo de datos, a lo largo de un proceso de trabajo con una capa de elevaciones: a) mapa escaneado, b) curvas de nivel, c) capa ráster, d) curvas de nivel.

altitud correspondientes. Se trata de una capa ráster, pero los análisis que podemos realizar a partir de ella son prácticamente nulos. En realidad, no es más que un «dibujo». Para obtener un dato más susceptible de análisis debemos vectorizar esas curvas de nivel, convirtiéndolas en una capa vectorial de líneas. Este proceso de vectorización ya se comentó en el capítulo 6, cuando vimos las fuentes de datos más habituales.

Como se dijo entonces, esa digitalización puede llevarse a cabo manualmente, bien digitalizando en pantalla o bien con alguna otra herramienta más específica. Esa conversión no es, sin embargo, la que interesa a este capítulo, ya que aquí veremos los procesos y algoritmos que pueden automatizar esta tarea de reconocer las líneas de ese mapa escaneado y delinear en base a ellas las entidades de la capa vectorial resultante.

A partir de esa capa de líneas, el siguiente paso es la obtención de una capa ráster. Todo lo necesario para llevar a cabo este proceso ya lo hemos visto en el capítulo 14.

Por último, a partir de esa capa ráster podemos obtener de nuevo un dato vectorial tal como un TIN, o bien de nuevo unas curvas de nivel. En el caso de la figura, las curvas de nivel son anteriores a la capa ráster, con lo cual este paso no resulta de gran utilidad, pues ya disponemos de ellas. Sin embargo, sabemos que hay tecnologías que permiten obtener una capa de elevaciones ráster sin necesidad de partir de un dato vectorial, y en ese caso la obtención de este último implica necesariamente un cálculo a partir de la

capa ráster. Las curvas de nivel pueden sernos útiles para, por ejemplo, generar un resultado cartográfico, y en este capítulo veremos en qué se basan las formulaciones que nos permiten hacer tal conversión.

A lo largo de este capítulo vamos a ver dos clases de procesos para crear capas vectoriales a partir de capas ráster.

- Conversión de capas ráster discretas. Vectorización.
- Conversión de capas ráster continuas.

Partiendo de una capa ráster discreta, el proceso de vectorización pretende generar una capa vectorial que recoja las geometrías que aparecen en dicha capa. Estas geometrías vienen definidas por la forma en que los valores de las mismas clases se disponen en la malla de celdas. Este es el caso que encontramos cuando disponemos de una capa ráster pero el modelo conceptual del espacio geográfico no es el modelo de campos, sino un modelo de entidades discretas. Cada una de estas entidades se constituyen mediante conjuntos de celdas contiguas con el mismo valor. Esta idea se recoge en la figura 18.2

En la conversión de capas continuas, veremos cómo expresar una capa ráster continua mediante un conjunto de entidades, en particular mediante líneas y polígonos. Para el caso de líneas, analizaremos cómo delinear un conjunto de isolíneas a partir de una capa ráster. En el caso de polígonos, estudiaremos cómo crear un TIN, estructura de triángulos irregulares que, como ya vimos, sirve para almacenar variables continuas, particularmente la elevación.

18.2. Vectorización de entidades

La vectorización de entidades tiene como base una capa ráster con una variable de tipo nominal u ordinal, en la cual se reflejan distintas categorías. El objetivo es expresar mediante entidades vectoriales los conjuntos de celdas de una misma categoría.

La necesidad de efectuar este tipo de operaciones va a aparecer habitualmente cuando encontremos en nuestro trabajo una capa ráster de este tipo, pues ya sabemos que resulta más adecuado en la mayoría de casos almacenar esta información según el modelo de datos vectorial.

Capas tales pueden surgir en el trabajo con SIG en diversos momentos, pues muchas operaciones de análisis sobre capas ráster van a generar este tipo de resultados. La conversión de dichos resultados al modelo de datos vectorial va a permitir darles un uso distinto y combinar ambos modelos de representación, utilizando cada cual cuando más convenga.

Estas operaciones pueden partir de capas ráster con variables continuas, de las que se extraen entidades de acuerdo a formulaciones diversas. Un ejemplo muy característico es la delineación de redes de drenaje y la delimitación de cuencas vertientes, operaciones ambas que se realizan a partir de un Modelo Digital de Elevaciones, como vimos en el capítulo 16.

Otros procesos de vectorización que ya conocemos son los que se llevan a cabo a partir de cartografía escaneada. En este caso, no obstante, la situación es bien distinta, ya que lo que a simple vista parece una misma línea o un mismo polígono en el mapa escaneado, realmente no es un conjunto de celdas con un único valor (es decir de un único color), sino con varios valores (colores) similares. Esta situación hace más difícil trabajar con este tipo de capas a la hora de vectorizar y reconocer las entidades que se deben vectorizar, y requiere procesos previos de tratamiento para que ese mapa escaneado se encuentre en las mejores condiciones antes de proceder a la vectorización. La figura 18.3 muestra una imagen

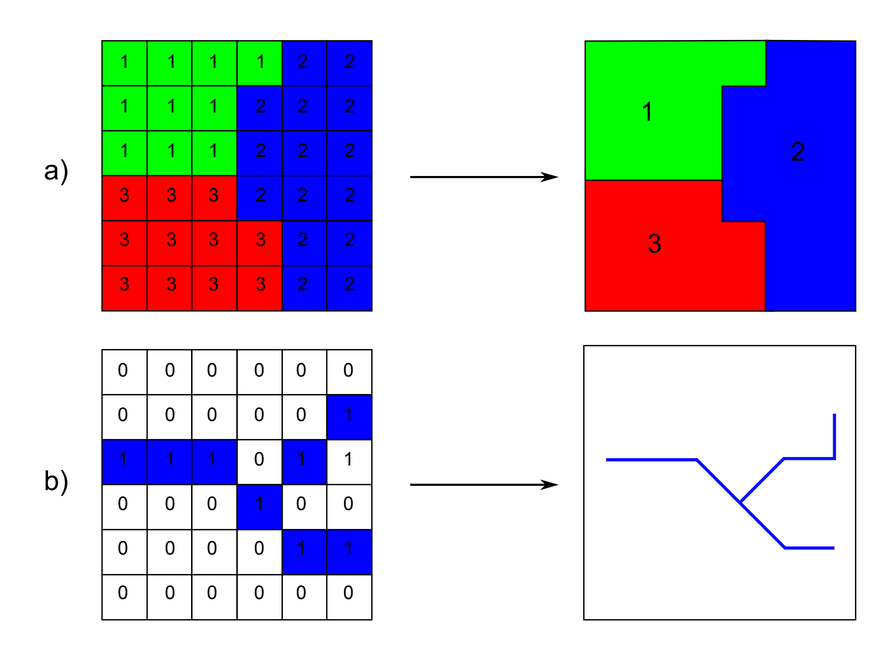


Figura 18.2: Vectorización de capas ráster discretas en capas de polígonos (a) y de líneas (b). .
.

que contiene líneas y polígonos, y junto a ellas las representaciones de estas que resultan óptimas para proceder a vectorizarlas como capas de líneas o capas de polígonos respectivamente.

El paso de las versiones originales a estas versiones óptimas implica el uso de operaciones de álgebra de mapas, así como las que se vieron en las secciones 17.3.1 o 17.4.2, dentro del capítulo dedicado al tratamiento de imágenes. Los algoritmos que trataremos en esta sección supondrán ya que se trabaja sobre una capa donde las entidades a vectorizar están representadas por valores uniformes, dependiendo su rendimiento de esta circunstancia. Aun así, haremos mención también a los procesos de preparación previos, ya que en algunos casos son parte inseparable de ciertos procesos de vectorización muy frecuentes en un SIG.

Algoritmos para el tratamiento de imágenes no necesariamente en estas circunstancias óptimas se detallan en [250], donde pueden encontrarse algoritmos para vectorización de todo tipo de imágenes en color. Estos algoritmos están enfocados a un trabajo fuera del ámbito de los SIG, y por ello pueden no satisfacer exactamente las necesidades que se presentan dentro de nuestro campo.

18.2.1. Vectorización de líneas

Volviendo a los algoritmos sobre capas discretas donde las entidades a definir están claramente definidas, comencemos con los de vectorización de líneas. Estos son similares en cierta medida a los de polígonos que veremos más adelante, aunque entrañan en principio algo menos de dificultad técnica.

Una característica reseñable de los algoritmos de vectorización de líneas son sus mayores requerimientos en lo que respecta a las características de la capa de partida. Esto es así debido

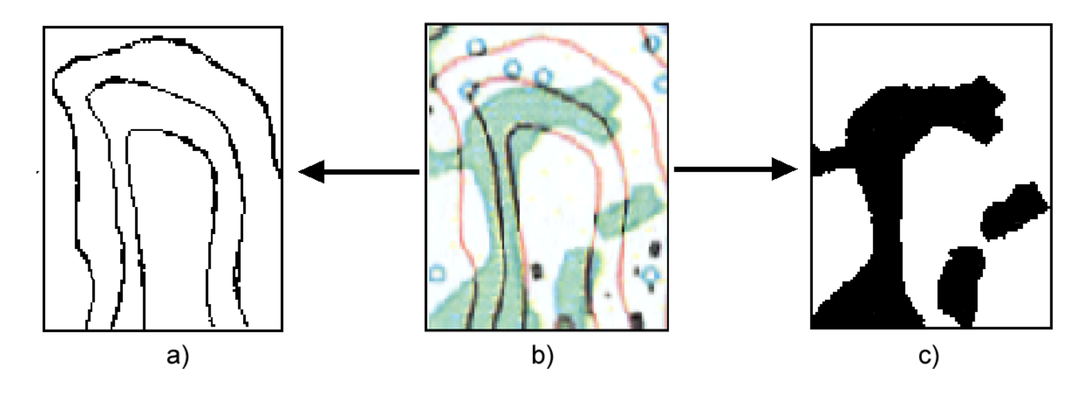


Figura 18.3: b)Imagen con polígonos y líneas para vectorizar. a) y c) La misma imagen tras ser trasformada con objeto de mejorar las condiciones para la vectorización (como capas de líneas y polígonos respectivamente) y aumentar la calidad de esta.

a que, a la hora de vectorizar una línea, vamos a necesitar que esta sea, en efecto, una línea como tal. Es decir, que el conjunto de celdas que la definen sea un conjunto de un único píxel de ancho, tal y como vimos en el apartado 17.4.2 al tratar las operaciones de esqueletización.

Para el caso de polígonos, si en la obtención de una imagen binaria los polígonos han quedado más grandes o más pequeños, o con una forma u otra, este hecho no afecta al resultado. El algoritmo los reconocerá y los vectorizará, y su funcionamiento será igual en todos casos. Las imprecisiones que puedan existir en la imagen binaria se trasladarán a las capas vectoriales resultantes, pero los algoritmos de vectorización no verán afectado su funcionamiento por dichas imprecisiones.

En el caso de las líneas encontramos una situación bien distinta. La presencia de píxeles adicionales o la falta de conectividad entre líneas van a dificultar seriamente la vectorización, ya que los algoritmos correspondientes están pensados para trabajar sobre entidades lineales que cumplan las características citadas anteriormente.

En ocasiones, las capas con las que trabajamos van a estar en buenas condiciones de ser vectorizadas (por ejemplo, en el caso ya citado de una red de drenaje extraída a partir de un MDE). En otras muchas, sin embargo, las capas ráster con las que vamos a trabajar van a requerir procesos previos, habitualmente laboriosos. Por esta razón, la obtención de una capa óptima a vectorizar es el obstáculo principal que encontraremos, por ejemplo, cuando trabajemos con imágenes, ya que estas en la mayoría de los casos no cumplen las condiciones requeridas y hacen necesario un importante trabajo de preparación.

Para ver un ejemplo de lo anterior, trataremos una de las tareas que con más frecuencia se presenta dentro del trabajo con SIG: la digitalización de curvas de nivel de un mapa impreso. Automatizar la vectorización en sí no es en este caso una tarea altamente compleja, pero los procesos tales como la segmentación o las operaciones morfológicas que ya conocemos no resultan sencillos de ajustar de forma automática para obtener un resultado vectorizable. Es necesario en la mayoría de los casos un cierto grado de trabajo manual.

Debe pensarse que el mapa escaneado es habitualmente una imagen en color, más compleja que las imágenes binarias sobre las que se presentaron en el capítulo 17 las operaciones morfológicas. Por ello, resulta más difícil ajustar estos procesos.

El paso de una imagen original tal como el mapa escaneado de la figura 18.3 a uno más apto para la vectorización como el de la imagen a) de dicha figura es un proceso más complejo de lo que puede parecer, ya que comporta las siguientes dificultades:

- Las líneas tienen más de un píxel de ancho
- Las líneas tienen píxeles de varios colores distintos
- Aparecen elementos adicionales que dificultan la identificación automática de las líneas, pues rompen su continuidad. Por ejemplo, es habitual que las líneas se encuentren interrumpidas por etiquetas en las que se refleja su valor de elevación correspondiente. Otros elementos tales como ríos o carreteras pueden estar representados en el mismo mapa, y trazarse por encima de las curvas de nivel, que suelen quedar en un segundo plano. Mientras que estos últimos se trazan con colores diferentes a las curvas de nivel, las etiquetas tienen el problema añadido de que presentan el mismo color que la propia línea, con lo cual no pueden eliminarse en base a dicho color con procesos tales como la segmentación (véase el apartado 17.3.1) [251].

El problema más grave de los anteriores es la falta de continuidad, el cual puede tratar de solucionarse tanto en la parte ráster, como en la vectorial, es decir, antes o después de la vectorización como tal. Para el ojo humano, resulta sencillo en caso de discontinuidad intuir cual es la conectividad de las lineas, ya que nuestra percepción tiende a agrupar aquellos elementos que sugieren una continuidad lineal, considerándolos como una entidad única pese a que en realidad no se encuentren unidos y sean objetos aislados. Por desgracia, un SIG no comparte con nosotros estas capacidades perceptivas, y es necesario que las uniones entre los tramos de líneas existan como tales de un modo más tangible.

Para el caso ráster, ya vimos en su momento (apartado 17.4.2) las operaciones morfológicas que nos pueden ayudar a «conectar» las líneas cuando hayan quedado separadas por alguna razón. Para el caso vectorial, existen de igual forma diversos algoritmos que tratan de realizar esta conexión una vez que la vectorización se ha producido y aparecen imprecisiones. En [252] puede encontrarse uno de tales algoritmos vectoriales, que no se detallarán aquí por estar fuera al alcance de este capítulo.

En líneas generales, y teniendo en cuenta los condicionantes anteriores, la vectorización de curvas de nivel a partir de mapas topográficos se lleva a cabo siguiendo una cadena de procesos que comporta los siguientes pasos.

- Escaneado del mapa original
- Filtrados y mejoras sobre la imagen resultante
- Segmentación
- Operaciones morfológicas
- Conversión ráster–vectorial

siendo en este último paso en el que nos centramos dentro de este capítulo.

Para esta conversión ráster–vectorial, el algoritmo correspondiente debe seguir un esquema genérico como el siguiente, a aplicar para cada una de las curvas de nivel.

- Localizar una celda de la curva.
- Analizar las celdas alrededor suyo para localizar la siguiente celda de la curva.
- Desplazarse a esa celda.
- Repetir los pasos anteriores.

Este proceso se detiene hasta que se cumple un criterio dado, que es el que concluye la vectorización de la línea. En general, el criterio a aplicar

suele ser que se regrese a la celda de partida, o que no pueda encontrarse una nueva celda entre las circundantes (a excepción de la anterior desde la cual se ha llegado a esta).

Siguiendo este esquema, se dota de un orden a las celdas de la línea. Mediante esta secuencia ordenada de celdas, puede crearse la línea vectorial como una secuencia ordenada de coordenadas (particularmente, las coordenadas de los centros de esas celdas).

Hay muchos algoritmos distintos de vectorización. Las diferencias van desde la forma en que se analizan las celdas circundantes o el criterio que hace que se concluya la vectorización de la línea, hasta formulaciones más complejas que siguen un esquema distinto.

Más allá de lo visto en este apartado, la conversión de mapas topográficos en capas de curvas de nivel es un proceso complejo del que existe abundante literatura. Buenas visiones generales de este pueden encontrarse en [253, 254, 255, 256]

Además de la vectorización de curvas de nivel, tarea habitual que ya hemos visto, existen otras muy variadas que presentan cada una de ellas distintas circunstancias. Así, y aún disponiendo ya de una capa que presente las condiciones idóneas para ser vectorizada, el proceso puede presentar más dificultad de la que con lo visto hasta este punto puede pensarse. No hay que olvidar que, a la hora de vectorizar un conjunto de líneas, estas se definen no únicamente por su forma, sino por otros elementos tales como, por ejemplo, la topología.

En el caso particular de vectorizar un conjunto de curvas de nivel tal y como lo venimos detallando, esta topología es clara y no entraña dificultad añadida a la hora de vectorizar. Por propia definición, las curvas de nivel no pueden cruzarse con otras curvas, con lo cual basta seguir el contorno de las mismas y no preocuparse de estos cruces. Sin embargo, vectorizar otros elementos implica tener en cuenta circunstancias distintas.

Algo más complejo que vectorizar curvas de nivel es hacerlo, por ejemplo, con una red de drenaje (recordemos que en el capítulo 16 vimos cómo extraer esta en formato ráster a partir de un MDT). En este caso sí que existen intersecciones, pero, puesto que los cauces solo tienen un único sentido, las formas que el conjunto de estos puede adquirir están limitadas. En particular, una red de drenaje es siempre una estructura en árbol, lo cual quiere decir que no van a existir rutas cíclicas en dicha red[1].

Si lo que tratamos de vectorizar es una red viaria, las posibilidades son más amplias, y además, como ya sabemos, el modelo ráster no es adecuado para registrar completamente la topología de dicha red. La existencia de rutas cíclicas complica además los algoritmos de vectorización en caso de que estos pretendan añadir topología a sus resultados, con lo que la operación no resulta tan sencilla como en el caso de las curvas de nivel.

En resumen, la existencia de topología añade complejidad a la vectorización de líneas. Aún así, siempre resulta posible (y en muchos casos suficiente) vectorizar estas y obtener como resultado una capa sin topología (*spaguetti*). Esta capa, ya en el modelo vectorial, puede tratarse posteriormente para dotarla de la topología necesaria, en caso de que así se requiera.

18.2.2. Vectorización de polígonos

Muy relacionada con la vectorización de líneas, tenemos como herramienta habitual dentro de un SIG la vectorización de polígonos. Esta guarda gran similitud con la anterior, en cuanto que el proceso se basa también, fundamentalmen-

[1]Desde el punto de vista de la teoría de grafos, un árbol es un grafo *acíclico conexo*.

te, en ir siguiendo una serie de puntos y conectarlos, para de este modo definir el contorno del polígono.

Los puntos que nos interesan para la delineación de un polígono no son, a diferencia del caso de líneas, todos los que conforman el objeto a vectorizar, sino tan solo una parte de ellos. En concreto, van a resultar de interés únicamente las celdas exteriores, es decir, las que al menos tienen una celda circundante con un valor distinto al del propio polígono a vectorizar (en el caso de trabajar con una imagen binaria, las que tienen al menos una celda circundante que pertenece al fondo). Las celdas internas no nos aportan información relevante, ya en ellas no se va a situar ninguna de las coordenadas de la entidad vectorial que buscamos.

Siendo dichas celdas externas las que debemos tratar para delinear la entidad vectorial, el primer paso es, por tanto, su localización. Esta no es difícil si recordamos algunas de las operaciones morfológicas que vimos en el capítulo 17. Por supuesto, la aplicación de estas exige que se den, una vez más, una buenas condiciones en la imagen, en particular que la frontera del polígono a vectorizar esté claramente definida. Para ello basta, como ya hemos dicho, que el valor dentro del polígono sea uniforme (valor que consideraremos como si fuera el 255 o 1 de las imágenes binarias, según lo comentado en su momento), y distinto de los restantes a su alrededor (valores todos ellos que consideraremos como el valor 0 de las imágenes binarias, que entonces identificábamos con el fondo)

De las operaciones morfológicas que conocemos, la erosión nos da una idea de la forma de proceder a la hora de localizar las celdas importantes. De la forma en que lo presentamos entonces, el proceso de erosión elimina aquellas celdas que se sitúan en contacto con el fondo y están en el borde del objeto. Estas son exactamente las que nos interesan de cara a la vectorización. Como muestra la figura 18.4, la diferencia entre una imagen binaria y dicha imagen tras un proceso de erosión es el contorno del objeto.

El proceso de erosión se aplica en este caso con un elemento estructural como el siguiente, en lugar del que vimos en el apartado 17.4.2.

0	1	0
1	1	1
0	1	0

Esto no es estrictamente necesario ya que, de aplicar el que vimos entonces, y aunque el contorno resultante sería distinto, el resultado sería igualmente vectorizable[2].

Sobre este contorno, el proceso de digitalización ya no difiere, a primera vista, del que efectuábamos sobre una curva de nivel. El objetivo es asignar un orden a las celdas de ese contorno, de modo que siguiendo dicho orden quede definido el perímetro del polígono.

[2]Otra forma de ver esto es la siguiente. A la hora de ir recorriendo el contorno podemos hacerlo desplazándonos solo en horizontal y vertical (como lo haría la torre en un tablero de ajedrez), o bien en todas direcciones (como lo haría la reina). Con el elemento estructural que vimos en la sección 17.4.2 el contorno resultante puede ser recorrido haciendo movimientos de torre, mientras que con el que hemos aplicado en este caso (y como puede apreciarse en la figura) recorrer el contorno exige movimientos en diagonal.
En la jerga del tratamiento de imágenes se dice que existe *conectividad-4* en el primer caso y *conectividad-8* en el segundo, haciendo referencia al numero de celdas circundantes a las que uno puede desplazarse según cada uno de estos esquemas de movimiento
Otra forma habitual de referirse a estos conceptos es hablar de *vecindad de Von Neumann* para el caso de 4 posibles conexiones o *vecindad de Moore* para el caso de 8.

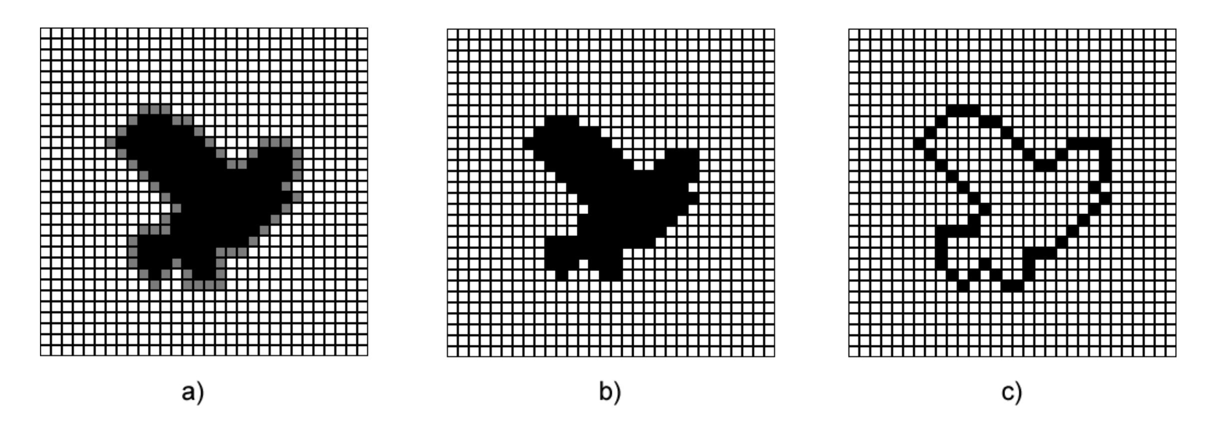


Figura 18.4: a) Imagen base con un polígono a vectorizar. Las celdas que forman el contorno se muestran en gris. b) La imagen anterior tras un proceso de dilatación. c) Contorno del polígono obtenido a partir de la diferencia entre las dos imágenes anteriores.

Para ello, y como en el caso de las curvas de nivel, basta comenzar en uno de los puntos e ir siguiendo de un modo sistemático el contorno, añadiendo las coordenadas de los puntos recorridos. Dichas coordenadas, no obstante, no son en este caso las de los centros de las celdas, sino que se deben tomar las de los vértices para que esta forma se almacene el contorno de cada una de las celdas externas al objeto vectorizado. Particularmente, son de interés las coordenadas de aquellos vértices que se sitúan en el lado exterior del contorno. Esto puede comprenderse mejor viendo la figura 18.5.

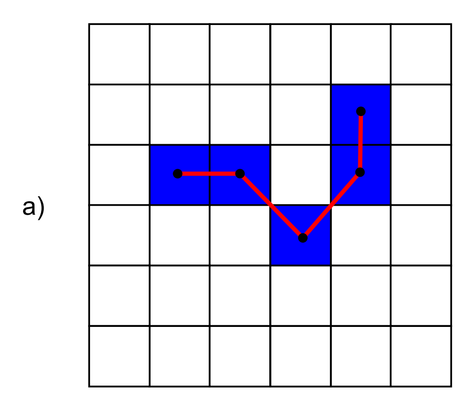
Los resultados en los dos supuestos representados en la figura son bien distintos, a pesar de que el objeto a vectorizar es el mismo, pero en un caso se interpreta como una línea y en otro como un polígono.

Un algoritmo para vectorizar el contorno de un polígono debe localizar una celda de dicho contorno e ir avanzando hasta rodear por completo este, almacenando las coordenadas de los bordes exteriores de todas las celdas recorridas. El avance se detiene cuando se vuelve a la misma celda en la que se comenzó, momento en el cual el proceso de vectorización queda completo. En [257] puede encontrarse con más detalle la descripción de un algoritmo de esta forma, parte de cuyos fundamentos pueden emplearse igualmente para la vectorización de líneas.

Al igual que sucedía con dichas líneas, la complejidad del objeto puede presentar problemas para su vectorización. El algoritmo anterior no contempla, por ejemplo, la presencia de «huecos» en el polígono. Como ya conocemos, esos huecos son polígonos internos que han de incorporarse a su vez a la entidad, y por tanto deben ser a su vez vectorizados.

18.3. Isolíneas

Otro resultado vectorial que puede generarse a partir de una capa ráster son las *isolíneas* o curvas de nivel. Aunque el caso más habitual es utilizarlas para representar elevaciones (tal y como aparecen en un mapa topográfico habitual), se emplean también para otro tipo de variables de tipo continuo tales como la presión (*isobaras*), la precipitación(*isoyetas*) o el tiempo(*isocronas*).



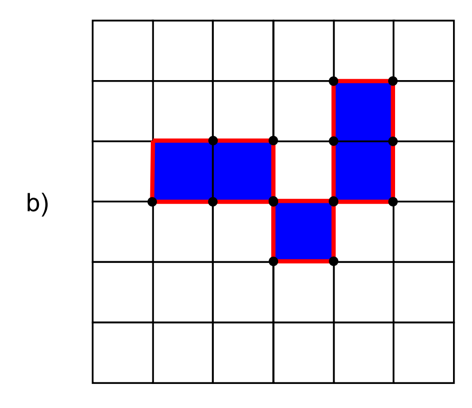


Figura 18.5: Mientras que en la vectorización de líneas (a) se toman las coordenadas del centro de la celda, en el caso de polígonos (b) se deben utilizar las de los vértices para delinear completo el contorno del objeto.

La forma de obtenerlas a partir de una capa ráster es, no obstante, la misma en todos los casos.

En un apartado anterior, veíamos cómo vectorizar un mapa escaneado que contiene el trazado de dichas curvas de nivel. En este, veremos cómo hacerlo cuando disponemos de los datos de elevación como tales en lugar de unas curvas ya trazadas en una imagen. Mientras que en aquel caso obteníamos curvas de nivel con la equidistancia utilizada en el mapa escaneado (es decir, el trabajo de traza las curvas ya estaba hecho y el proceso no hacía sino convertirlas en entidades vectoriales), aquí podemos definir nosotros esa equidistancia y los valores para los que queremos obtener tales curvas.

El calculo de curvas de nivel puede realizarse a partir de una capa ráster, pero también a partir de una capa de puntos con datos de elevación. En este ultimo caso, no obstante, ya sabemos que podemos convertir esos puntos en una capa ráster mediante métodos de interpolación (según lo visto en el capítulo 14), y después en base a este calcular las curvas de nivel. En esta sección trataremos únicamente la delineación de curvas de nivel a partir de capas ráster.

Un método basado en triangulación para obtener curvas de nivel a partir de puntos distribuidos irregularmente se detalla en [258]. Una revisión detallada de métodos disponibles para esta tarea se puede encontrar en [259].

Respecto al cálculo a partir de una estructura regular como una capa ráster, los algoritmos correspondientes no derivan únicamente del trabajo con SIG, sino que se trata de un área muy desarrollada en el tratamiento de imágenes digitales. Las curvas de nivel ponen de manifiesto las transiciones existentes en los valores de la imagen, y estos puede resultar de interés para una mejor interpretación de esta o la automatización de ciertas tareas.

Existen dos enfoques principales a la hora de trazar curvas de nivel en base a una malla de datos regulares [260]

- Seguimiento de líneas
- Análisis por celdas

En el seguimiento de líneas, se localiza un punto que pertenezca a la curva de nivel y después se «sigue» esa curva de nivel hasta que se alcanza un borde de la malla ráster, o bien la curva se cierra regresando al punto inicial.

La localización de un punto de la curva se hace empleando métodos de interpolación, del mismo modo que veíamos en la sección 15.3 al tratar las técnicas de remuestreo. Como es lógico pensar, los valores de las curvas de nivel que se van a trazar son en generales valores «redondos» (múltiplos de 100 o 200, por ejemplo). Por el contrario, los valores de las celdas pueden ser cualesquiera, y salvo en casos particulares, los valores de las curvas de nivel no van a presentarse en los valores exactos de las celdas, que por convenio corresponden como sabemos a los centros de estas.

Dicho de otro modo, las curvas de nivel no pasan por los centros de las celdas, pero sí que atraviesan estas. Los métodos de interpolación se emplean para saber por dónde atraviesa exactamente la celda la curva de nivel correspondiente, ya que con los valores de una celda y sus circundantes, referidos a sus respectivos centros, pueden estimarse valores en puntos no centrales de estas.

Cuando una curva de nivel entra en una celda, obligatoriamente debe salir de ella (una curva de nivel, por definición, no puede concluir bruscamente salvo que se encuentre en el extremo de la malla de celdas). Los mismos métodos de interpolación se emplean para calcular por dónde sale y hacia qué celda lo hace. Sobre esa celda se aplicará un análisis similar, y de este modo se produce ese seguimiento de la línea que resulta en el trazado completo de la curva de nivel.

El resultado de este proceso es un conjunto de puntos que unidos secuencialmente conforman la curva de nivel buscada.

Estos métodos tienen la ventaja de que, al presentar la línea como un continuo, dan resultados mejores para su representación (esto era especialmente importante cuando se empleaban *plotters* para la impresión de esos resultados), y es más fácil etiquetar el conjunto de líneas [261]. Esto es así debido a que los métodos que realizan un análisis por celdas no tratan la curva de nivel como una única entidad, sino como un conjunto de pequeños tramos, cada uno de los cuales definido en el interior de una única celda.

Aunque el resultado visualmente puede ser el mismo, la capa generada mediante un método de seguimiento de curvas va a contener menos entidades y ser más correcta desde un punto de vista semántico, ya que una curva se expresa como una única entidad, no como un conjunto de ellas.

Entre los métodos de análisis por celdas, uno con gran relevancia (especialmente en el tratamiento digital de imágenes) es el conocido como *Marching Squares*, una adaptación bidimensional del algoritmo tridimensional *Marching Cubes* presentado en [262].

El fundamento de este método es el hecho de que, si una curva de nivel atraviesa una celda, existen únicamente 16 posibles configuraciones de los vértices de esa celda en función de si su valores correspondientes están *dentro* o *fuera* de la curva de nivel. La figura 18.6 muestra esas configuraciones.

En base a esto, se recorren todas las celdas, se analiza en cuál de las configuraciones se encuentra cada una de ellas y, cuando corresponda, se traza una línea entre el punto de entrada y de salida de la curva de nivel en la celda. Estos puntos se calculan, al igual que ya veíamos para los métodos de seguimiento, mediante interpolación.

El algoritmo no establece relación alguna entre las celdas por las que pasa una misma curva de nivel, sino que toma estas separadamente. Por ello, aunque visualmente los tramos que resultan estén conectados, son considerados como segmentos independientes a la hora de generarlos, y también a la hora de su manejo una vez la capa de curvas de nivel ha sido creada.

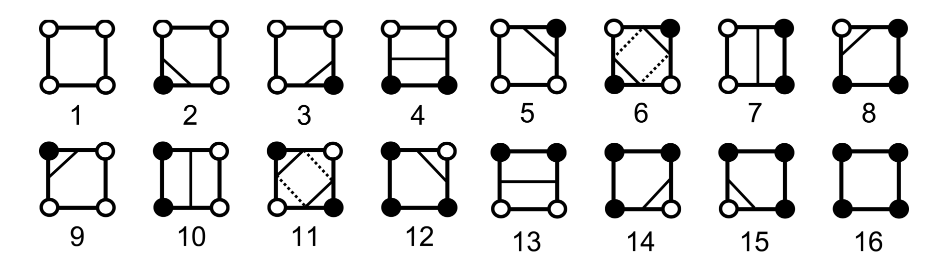


Figura 18.6: Posibles configuraciones de una celda según el algoritmo *Marching Squares*. Los vértices en negro estan dentro de la curva de nivel, mientras que los blancos se encuentran fuera.

.

Además de este inconveniente, el método presenta algunas ambigüedades, como puede verse en la figura 18.6 para los casos 6 y 11. En estos supuestos no está clara cuál es la forma en que la curva de nivel atraviesa la celda. En los métodos de seguimiento no existe esta deficiencia, ya que se sabe desde qué celda llega la curva de nivel.

18.4. Creación de TIN

Ya conocemos los TIN como estructuras vectoriales utilizadas para almacenar una variable continua tal como la elevación. Como vimos en el capítulo 5, estos están formados por un conjunto de puntos que, utilizados como vértices de triángulos, dan lugar a una red de triángulos interconectados que cubren toda la superficie estudiada.

Estos triángulos aparecen en mayor número en las zonas donde la variable presente una mayor variabilidad (relieve más abrupto) y en menor número cuando no existe una variabilidad tan acusada (relieve llano).

La creación de un TIN a partir de otro modelo de datos como, por ejemplo, una capa ráster, implica las siguientes operaciones:

- Selección de puntos más representativos.
- Triangulación de dichos puntos.

18.4.1. Selección de puntos

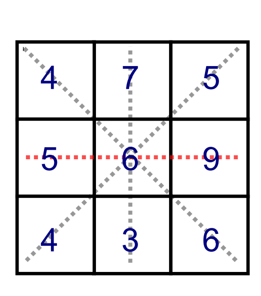
No todas las celdas de la capa ráster son igual de interesantes a efectos de crear un TIN. Si tomásemos todas las celdas y las utilizáramos como vértices, el TIN resultante sería sin duda preciso, pero el número de triángulos sería muy elevado. No estaríamos aprovechando la gran ventaja de los TIN, que no es otra que su adaptabilidad en función de la variabilidad de cada zona, ya que no estaríamos teniendo en cuenta esta variabilidad para tomar más o menos puntos según sea necesario.

Si se debe tomar un número de puntos menor, es necesario un método para eliminar aquellos que aportan menos valor al TIN resultante, de forma que, con los puntos que se consideren, este sea lo más preciso posible[3].

La selección de las celdas que deben considerarse como vértices de la red puede llevarse a cabo mediante diversos algoritmos. El algoritmo conocido como VIP(*Very Important Points*[4])[263] es uno de los más habituales. Se basa en evaluar la significancia de cada celda y después eliminar

[3]Elegir un número de puntos dado y obtener el mejor TIN posible con ese número no es una tarea en absoluto sencilla. Se trata de un problema matemático de los conocidos como *NP–Hard*.

[4]Puntos Muy Importantes



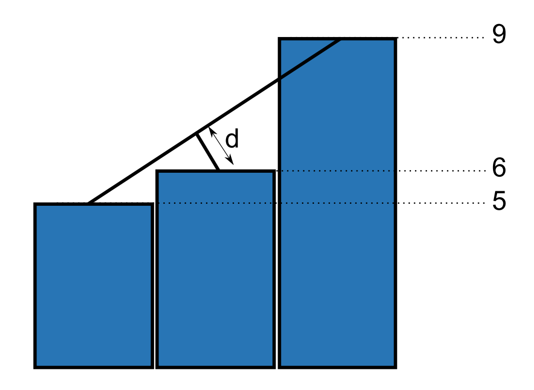


Figura 18.7: La significancia de una celda es la media de las significancias según las cuatro direcciones definidas. Para una dirección definida (en rojo), la significancia es la distancia d entre la celda central y la recta que une las celdas de los extremos

.

aquellas menos relevantes hasta quedarse con un número n de celdas, que serán las más adecuadas para formar el TIN. El criterio de eliminación de celdas menos significativas puede también basarse en un umbral de significancia, de forma que solo se consideren como vértices las celdas que lo superen.

La significancia de una celda se obtiene mediante un análisis local con una ventana 3×3, estudiando las cuatro posibles direcciones que pasan por la celda central (Figura 18.7).

En cada dirección se traza una recta que pasa por dos puntos extremos, teniendo en cuenta la elevación de los mismos. La distancia entre la celda central y dicha recta es la que define la significancia. La media de las cuatro significancias calculadas según todas las direcciones posible es la significancia global de la celda.

Otro algoritmo basado en análisis local es el propuesto por [264], que se basa en el análisis de las formas del terreno mediante ventanas 3×3 y 2×2, y busca hallar los puntos más representativos mediante la caracterización del tipo de forma del terreno. El análisis mediante la ventana 2×2 es muy similar a lo que vimos en el análisis hidrológico para la localización de celdas cóncavas, en el capítulo 16.

Un enfoque distinto a los anteriores es el propuesto por [265], denominado *drop heuristic*, que crea un TIN con todas las celdas[5] y después recorre iterativamente todos sus vértices, eliminando aquellos que se evalúen como de menor importancia.

18.4.2. Triangulación

Una vez que se tiene el conjunto de puntos significativos, es necesario conectar estos para formar la red de triángulos como tal, existiendo para ello existen diversas metodologías.

Para un conjunto de puntos V, una triangulación es una conjunto de triángulos que cumple las siguientes propiedades[266]:

- El conjunto de todos los vértices de esos triángulos es igual a V.
- Cada lado de un triángulo contiene únicamente dos vértices.

[5] Realmente no es un TIN *sensu stricto*, puesto que no es irregular. Los vértices presentan la misma estructura regular que caracteriza a la capa ráster de la que parte.

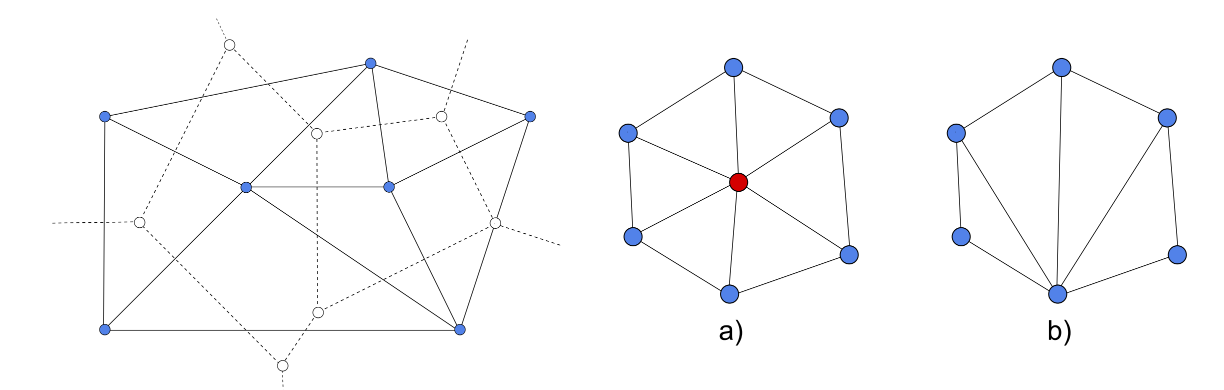


Figura 18.8: Teselación de Voronoi (en trazo discontinuo) y Triangulación de Delaunay (en trazo continuo) a partir de un conjunto de puntos (en azul)

.

Figura 18.9: Esquema de la eliminación de un vértice (en rojo) en una triangulación

.

- La intersección de dos triángulos cualesquiera es nula.
- La unión de todos los triángulos forma la envolvente mínima convexa del conjunto de puntos[6].

Los algoritmos de triangulación de un conjunto de puntos se basan la mayoría en considerar la tarea como un problema de optimización, siendo diversos los criterios aplicados. No obstante, es interesante que las propiedades de la triangulación guarden relación con el uso que vamos a darle al considerarla el elemento definitorio de un TIN, siendo esta una estructura en la que recogemos en general la forma de un terreno.

Puesto que dentro del triángulo van a asumirse unas propiedades constantes (pendiente, orientación), la idea es que los triángulos engloben áreas que, efectivamente, sean constantes en este aspecto. Esto se logra favoreciendo la creación de triángulos con ángulos cercanos a 60°, de modo que dichos triángulos sean lo más homogéneos posibles, evitándose aquellos de formas alargadas.

La forma más habitual de lograr esto es creando un tipo particular de triangulación conocida como *Triangulación de Delaunay*[267]. Esta presenta las siguientes propiedades:

- Dado un triángulo cualquiera de dicha triangulación, el círculo que en que se inscribe no contiene a ningún otro punto.
- Dados cuatro puntos que forman dos triángulos contiguos, la triangulación maximiza el mínimo ángulo interno del cuadrilatero formado.

Esta ultima propiedad es de especial interés para garantizar que los triángulos son óptimos en cuanto a su capacidad de representar fielmente el relieve.

La triangulación de Delaunay es el dual de otra estructura denominada *Teselación de Voronoi*, como puede verse en la figura 18.8. La teselación de Voronoi asocia a cada punto un polígono que representa el lugar geométrico de las coordenadas que tienen a dicho punto como punto más cercano de todos los del conjunto. Esto es similar a lo que veíamos para el método de interpolación

[6]Este es un polígono de gran importancia que veremos en la sección 19.6.1.

[7]Aunque en aquel momento lo analizábamos desde un punto de vista ráster.

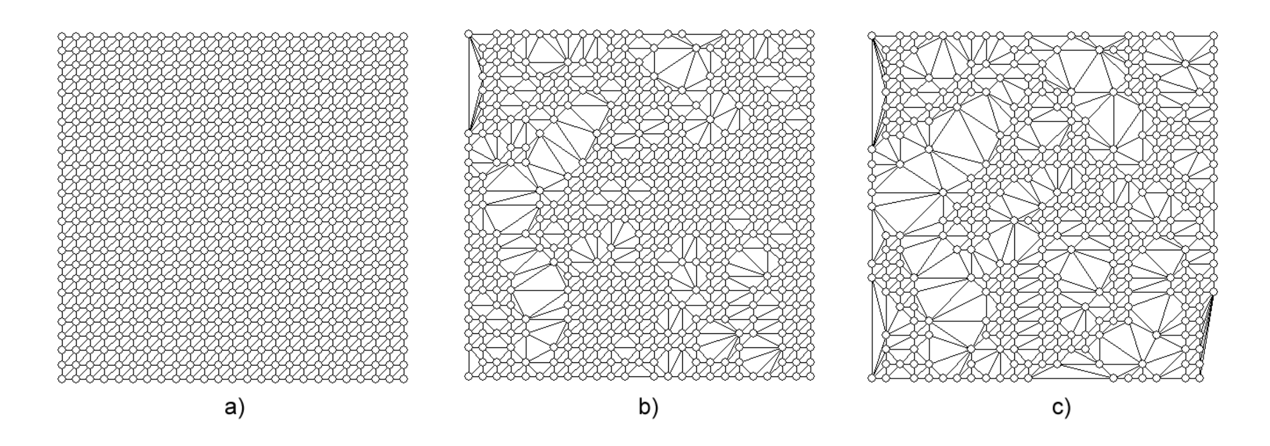


Figura 18.10: TIN original (a) y dos simplificaciones (b, c) en base al anterior.

.

por vecino más cercano, el cual genera, de hecho, una teselación de Voronoi[7].

Los algoritmos para crear una triangulación de Delaunay son abundantes, y existe mucha literatura al respecto. Comparaciones entre los más habituales y descripciones de estos pueden encontrarse en [268] y [269], donde se tratan con detalle los más importantes.

Aunque este tipo de triangulaciones son las más recomendables, la propia naturaleza de un TIN puede aprovecharse para crear otras triangulaciones derivadas que, aun no cumpliendo las condiciones de la triangulación de Delaunay, representan de forma más fiel el relieve. La incorporación de las denominadas *líneas de ruptura*[8] o *líneas de falla* es una de las modificaciones de mayor interés.

Dentro de un triángulo se asume que la pendiente y la orientación son constantes, y dicho triángulo define un plano en el espacio. Al pasar de un triángulo a otro, el cambio se produce justo en el lado que comparten dichos triángulos, representando dicho lado una línea de cambio. Si estas líneas de cambio coinciden con las líneas naturales en las que el relieve real que se quiere modelizar sufre un cambio brusco, el modelo obtenido será más fiel a la realidad.

En base a esta idea, puede forzarse a que dichas fallas naturales coincidan con los lados de los triángulos, definiéndolas explícitamente junto a los puntos que van a formar la base de la triangulación. La triangulación resultante no es de Delaunay, pero es más cercana a la verdadera forma del terreno, ya que incorpora información adicional.

18.4.3. Simplificación

Además de los métodos anteriores para seleccionar un conjunto reducido de puntos significativos y los algoritmos para obtener una red de triángulos a partir de estos, otro procedimiento importante es la simplificación de una red ya creada.

Como se muestra en la figura 18.9, pueden eliminarse puntos de una triangulación y rehacer esta con los puntos restantes. Este es el funda-

[8] *Breaklines*, en inglés

mento del proceso de simplificación, eliminando progresivamente vértices y, cada vez que uno de ellos es eliminado, recalculando la triangulación de la mejor forma posible.

Este procedimiento de simplificación es similar al que vimos para el algoritmo de selección de puntos *drop heuristic*, que partía de un TIN muy denso con todas las celdas ráster como vértices. Si en lugar de partir de dicho TIN se comienza con otro calculado según algún otro algoritmo, este puede irse simplificando hasta alcanzar un nuevo nivel de precisión fijado de antemano. La figura 18.10 muestra un TIN y dos versiones simplificadas del mismo, de distinto grado de simplificación. Nótese cómo en la imagen, si se eliminan puntos del exterior, varía el contorno de la triangulación.

Más detalle sobre algoritmos de simplificación puede encontrarse en [270].

18.5. Resumen

Las capas ráster pueden convertirse en capas vectoriales empleando algoritmos a tal efecto. Estos algoritmos pueden aplicarse sobre capas ráster de variables continuas o discretas, siendo distintos en cada caso. En el caso de variables discretas, las capas vectoriales generadas definen las geometrías que forman las distintas clases dentro de la capa ráster. Estas geometrías pueden ser poligonales o lineales.

En el caso de variables continuas, las capas vectoriales que se generan pueden ser de los tres tipos básicos de geometrías: puntos, líneas o polígonos. Para generar una capa de puntos, la conversión no requiere ningún algoritmo específico, pues basta calcular los valores de la capa ráster en las coordenadas de dichos puntos. Para representar una variable continua mediante líneas, se crea una capa de isolíneas. Por último, para el caso de polígonos, una estructura como un TIN puede generarse igualmente a partir de la capa ráster, seleccionando los puntos más importantes y después triangulándolos.

Capítulo 19

Operaciones geométricas con datos vectoriales

19.1. Introducción

El análisis de los datos ráster se basa fundamentalmente en el análisis de sus valores. Esto es debido a que las entidades geográficas asociadas a esos valores (las celdas) presentan una regularidad estructural propia del formato, de la cual no se puede derivar nueva información. Más aún, según vimos en el capítulo 15, resulta necesario homogeneizar dicha estructura, lo cual centra la atención del análisis en los valores recogidos y no en las formas geométricas a las que dichos valores caracterizan.

Con los datos vectoriales, la situación es diferente. En el caso de datos vectoriales, las entidades asociadas a cada atributo tienen sus propias características espaciales y la geometría que definen sirve por sí sola para llevar a cabo numerosos análisis. Si la unimos a los atributos que esta geometría lleva asociados, tenemos la posibilidad de realizar un número mayor de dichos análisis.

En este capítulo veremos una serie de operaciones que transforman los datos vectoriales actuando sobre sus geometrías, con el concurso en algunos casos de los atributos de estas. Los resultados de estas operaciones son nuevas capas cuyas geometrías aportan información adicional a las geometrías originales, o bien las transforman para que su uso sea más adecuado en otros análisis u operaciones.

Una buena parte de estas operaciones trabajan con dos capas, siendo similares en concepto (que no en la naturaleza de las operaciones que implican) a las funciones de tipo local del álgebra de mapas, empleadas para combinar varias capas con distintos valores. Otras requieren una sola capa, y aplican transformaciones en función de valores y geometrías o simplemente geometrías, pudiendo considerarse en cierta forma el equivalente a las funciones focales del álgebra de mapas ráster.

Muchas de las operaciones geométricas que pueden realizarse con datos vectoriales pueden llevarse a cabo empleando el álgebra de mapas de capas ráster si disponemos de la misma información en ese formato. Otras, sin embargo, no son adecuadas para efectuarse sobre una base ráster, del mismo modo que algunas operaciones del álgebra de mapas no se pueden llevar a cabo (o al menos no de una forma que resulte adecuada) utilizando datos vectoriales.

Esta equiparación entre el álgebra de mapas ráster y estas transformaciones geométricas (también una especie de álgebra de mapas vectorial, en cierta medida) puede plantear una disyuntiva desde el punto de vista del usuario de SIG que pretende analizar sus datos. Utilizar uno u otro modelo a la hora del análisis representa una elección igual que lo es el optar por una u otra forma de almacenamiento, vayamos o no a efectuar análisis sobre los datos almacenados. Sin embargo,

no son opciones mutuamente excluyentes, y lo ideal es conocer los puntos fuertes y débiles de cada planteamiento, para así combinar ambos de la mejor forma posible. Como ya sabemos, los SIG hoy en día no son puramente ráster o vectoriales, sino que incluyen herramientas con ambos planteamientos. Utilizarlas conjuntamente es la forma más idónea de aprovechar toda su potencia.

19.2. Zonas de influencia

Una de las transformaciones más importantes con capas vectoriales es la creación de zonas de influencia, también conocidas como *buffers*. Esta transformación puede llevarse a cabo con entidades de tipo punto, línea o polígono, y su resultado siempre es una nueva capa de polígonos. Las áreas cubiertas por estos polígonos reflejan las zonas de influencia de cada entidad, influencia que se considera la ejerce hasta una distancia dada. Pueden verse también de forma inversa, como una influencia recibida, de tal modo que todos los elementos dentro de la zona de influencia afectan a la entidad que la genera.

Por ejemplo, comenzando con un ejemplo de entidades puntuales, sea un conjunto de antenas de radio y una coordenada que representa la posición en la que nos encontramos. Si estas antenas tienen un alcance máximo (una dimensión máxima de su zona de influencia), es sencillo ver si estamos dentro de dicho alcance y recibimos la señal de alguna o varias de ellas. Para ello, basta crear zonas de influencia de los puntos donde se sitúan las antenas y después comprobar si el punto donde nos situamos esta dentro de alguna de dichas zonas. Para esto último, ya conocemos formulaciones específicas como las que vimos en 11.3.

Como se muestra en la figura 19.1, este análisis se puede realizar de dos formas distintas, según se considere la influencia ejercida por las distintas antenas o la recibida por el punto de análisis. En el caso a), se calculan las zonas de influencia correspondiente a cada una de las antenas, y posteriormente se comprueba si el punto analizado se encuentra en el interior de alguna de ellas. En el caso b), por el contrario, el planteamiento es opuesto. Tomando el punto donde nos encontramos, calculamos una zona de influencia alrededor de dicho punto, y después comprobamos si alguna de las antenas cae dentro de esta zona. En ambos casos, si esa comprobación tiene un resultado positivo nos encontramos en un punto de cobertura. Si el resultado es negativo, estamos fuera del alcance de las antenas.

El hecho de poder plantear este análisis de dos formas distintas es debido a que, según lo visto en 11.3, en un espacio métrico se tiene que, para dos puntos A y B, la distancia entre A y B (d_{AB}) es igual a la distancia entre B y A (d_{BA}). Pueden calcularse zonas de influencia basadas en otro tipo de medidas que no cumplen la anterior propiedad, aunque las operaciones correspondientes se llevan a cabo sobre capas en formato ráster, y las estudiaremos en el capítulo 20.

Calcular una zona de influencia de un punto es un procedimiento sumamente sencillo, ya que esta zona es simplemente un circulo centrado en el punto y con radio la distancia máxima de influencia. En el caso de considerar líneas en lugar de puntos, la operación es conceptualmente similar, aunque el algoritmo subyacente es notablemente más complejo. No obstante, la interpretación es idéntica.

En la figura 19.3 podemos ver cómo el trazado de un cauce se transforma en un área que engloba todos aquellos puntos que se sitúan a una distancia del cauce menor que el umbral establecido.

Podemos utilizar este tipo de zonas de influencia para, por ejemplo, estudiar qué indus-

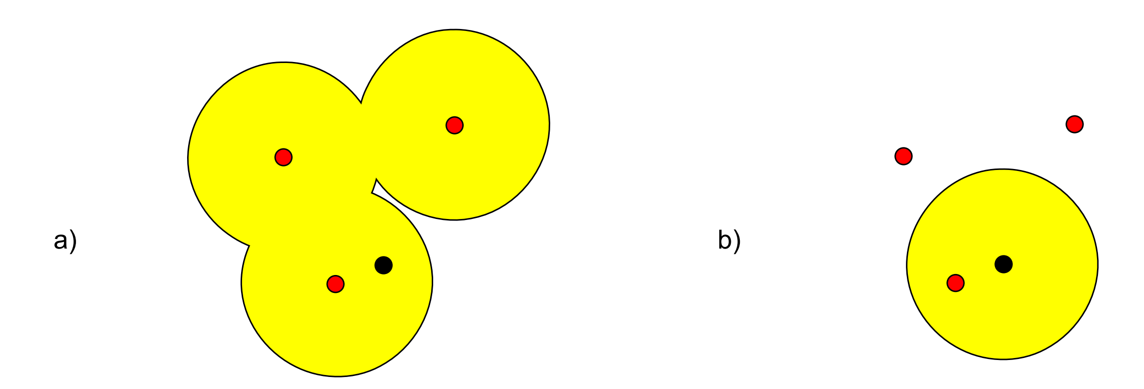


Figura 19.1: Zona de influencia a partir de puntos. La influencia puede analizarse a partir de los puntos que la ejercen (a) o los que la reciben (b)

trias pueden afectar la calidad de las aguas de un río. Teniendo el trazado de este río y el emplazamiento de las industrias, podemos hacer una zona de influencia del cauce para una distancia máxima de influencia y ver cuáles de estas últimas se sitúan dentro de ella.

De igual modo que en el ejemplo anterior, el análisis se pueden plantear de forma contraria, calculando las zonas de influencia de cada industria y viendo si estas cortan al trazado del cauce. Aquí podemos ver que, aunque los procedimientos para puntos y líneas sean distintos, su significado es el mismo, y por ello pueden utilizarse las zonas de influencia creadas a partir tanto de unas como de otras geometrías, indistintamente.

En el caso de polígonos, y puesto que ya conforman regiones, las zonas de influencia pueden establecerse no solo hacia el exterior, sino también hacia el interior de la geometría (Figura 19.2). En el caso de extender el área del polígono, la interpretación es la misma que en los ejemplos anteriores. Por ejemplo, para un polígono que delimita el perímetro de un núcleo urbano, su zona de influencia puede indicar la extensión de territorio al que los habitantes de ese núcleo podrían desplazarse en busca de algún servicio, o bien la de la región de procedencia de quienes pueden acudir a la ciudad en busca del mismo. Una zona de influencia hacia el interior, sin embargo, nos indicaría, por ejemplo, qué habitantes no es probable que salgan del núcleo urbano en busca de un servicio dado, ya que este desplazamiento es demasiado costoso.

Otro uso habitual de este tipo de zonas de influencia la encontramos en la gestión de costas. Si consideramos una zona de protección costera de una distancia dada, el polígono de influencia calculado hacia el interior con esa distancia representa las zonas a más de dicha distancia de la costa. Por tanto, representa a aquellas zonas fuera de la zona de protección, que serán aquellas sobre las que puedan establecerse actividades tales como la edificación o la implantación de infraestructuras.

Mientras que la zona de influencia exterior no tiene ningún límite, la trazada hacia el interior viene limitada por las propias dimensiones del polígono de origen.

Sobre la base anterior de zonas de influencia delimitadas a partir de las geometrías fundamentales, existen muchas variantes y formas de construirlas. Un proceso habitual es la creación no de una única zona de influencia para cada entidad base, sino de un conjunto de ellas equiespaciadas,

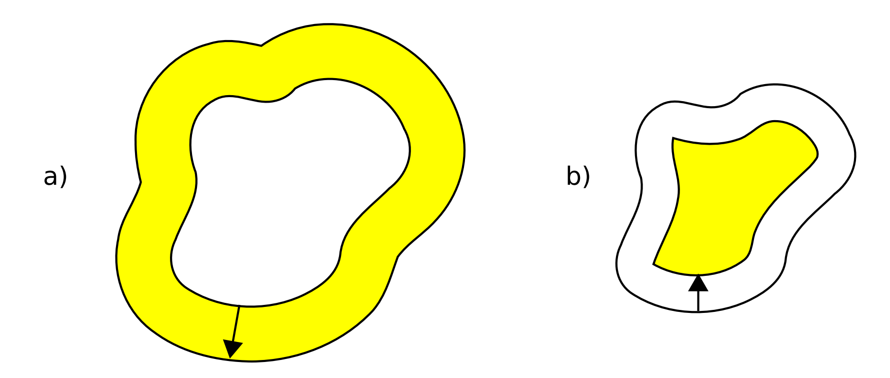


Figura 19.2: Zona de influencia a partir de polígonos. a) Hacia el exterior,b) hacia el interior.

de tal forma que puedan considerarse distintos grados de influencia. Este conjunto de zonas de tamaños sucesivos forman una serie de polígonos concéntricos que pueden, además, representarse de forma distinta para reflejar más explícitamente su significado.

Es fácil ver que si se trazan dos áreas de influencia de tamaño r y $2r$ a partir de una geometría de un tipo cualquiera, esta última es equivalente al área de influencia creada a partir de la primera, con distancia r.

La dimensión de una zona de influencia no debe ser constante para todos los elementos que componen la capa de base. Considerando el caso de una capa de cauces, no parece lógico considerar que todos los cauces van a tener una zona de influencia idéntica. Los cauces con un mayor caudal deberían tener una zona de influencia mayor, mientras que los pequeños afluentes deberían estar asociados a zonas de influencia menores.

Trabajando con una capa de puntos en lugar de una de líneas, como la mencionada en el ejemplo de las antenas, pudiera darse una situación en la que no todas ellas tuvieran la misma potencia y el alcance de su señal fuera distinto en cada caso. En este supuesto debería establecerse un radio distinto para los círculos de influencia de cada antena.

Tanto el caudal del cauce como la potencia de la antena son atributos de la entidad que estarán en su correspondiente tabla de atributos. Utilizando estos, pueden crearse capas de influencia variables que tengan en cuenta las características puntuales de cada elemento geográfico. Es decir, se toma en cada caso la dimensión del área de influencia de la tabla de atributos, en lugar de ser un valor fijo.

La figura 19.4 muestra el aspecto de una zona de influencia calculada a partir de una red de drenaje, cuya dimensión se establece en función del orden jerárquico de cada tramo.

Otra de las modificaciones que se pueden plantear a la creación de zonas de influencia es la simplificación de estas cuando existen solapes. En el resultado del caso a) la figura 19.1 puede verse cómo las distintas zonas de influencia se solapan en algunos puntos. No obstante la zona global que se ve afectada por la presencia de los puntos (antenas según suponíamos para el ejemplo), puede recogerse de igual modo con un único polígono, ya que, puesto que existe contacto entre las zonas, puede simplemente considerarse el contorno del conjunto.

En general, todas las entidades geográficas se prestan a la creación de zonas de influencia sin ninguna consideración especial. No obstan-

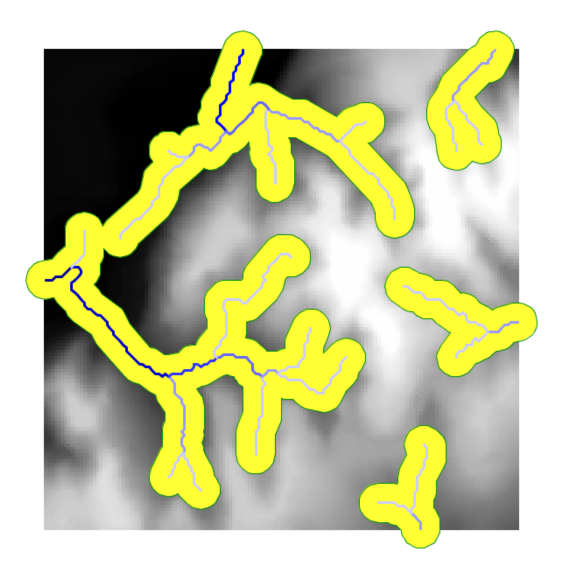


Figura 19.3: Zona de influencia a partir de elementos lineales (cauces).

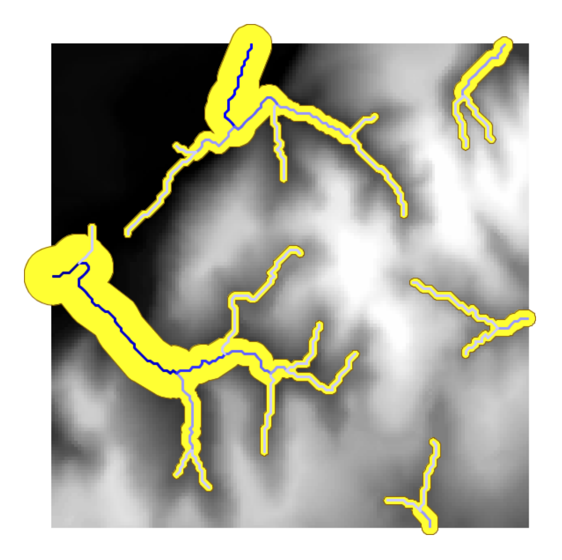


Figura 19.4: Zonas de influencia variable establecidas en función de los ordenes de los cauces.

te, en el caso de curvas o polígonos muy convolucionados, pueden en ocasiones darse formas incoherentes.

19.3. Operaciones de solape

Las operaciones de solape son equivalentes a las expresadas en el álgebra de mapas mediante funciones locales con varias capas ráster. Estas operaciones permiten generar nuevas capas vectoriales a partir del cruce de dos de ellas, pudiendo dichas capas de origen contener distintos tipos de entidades, aunque principalmente regiones (polígonos).

La naturaleza de estas operaciones es bien distinta a la de sus equivalentes ráster. Por una parte, se basan en cálculos geométricos que utilizan las coordenadas de cada entidad para obtener nuevas entidades resultantes. Por otra, se aplican operaciones lógicas sobre las capas de entrada para generar los resultados. Estas operaciones de tipo lógico (verdadero/falso), que como vimos también pueden aplicarse celda a celda entre dos capas ráster, son las que definen cada una de las operaciones de solape.

En dichas operaciones se aplican los conceptos de relaciones espaciales vistos en 11.4, y es en función de ellos como, a partir de los resultados geométricos, se generan las nuevas capas.

Encontramos los siguientes tipos de operaciones de solape:

- Recorte
- Diferencia
- Intersección
- Unión

19.3.1. Recorte

Una de las aplicaciones que veíamos de las funciones locales para capas ráster era la combinación con objeto de restringir la extensión de una capa de entrada. Utilizábamos una capa con la información de interés y otra con una «máscara», la cual indicaba qué celdas resultaba de interés preservar en la capa resultante.

En el caso vectorial, la operación de recorte toma una capa con cualquier tipo de entidades donde se contienen la información a recortar, y otra capa de polígonos que contiene aquellas regiones que resultan de interés. La capa resultante mantiene el mismo tipo de información, pero solo mantiene aquellas entidades que se incluyen total o parcialmente dentro de alguno de los polígonos de recorte, modificando dichas entidades cuando corresponda.

Esta operación se conoce como *clipping* en inglés y es habitual verla así citada o implementada en los SIG.

En la figura 19.5 podemos ver cómo la capa de entrada con polígonos, líneas y puntos se restringe a una extensión menor manteniendo dentro de dicha zona la misma información original. Se considera en este ejemplo un único polígono de recorte, pero pueden ser varios, e incluso polígonos con huecos interiores, sin que exista diferencia alguna en la operación.

Para los puntos, solo se mantienen aquellos que se sitúan dentro del polígono de recorte. Para las líneas, se mantienen aquellas que entran dentro del polígono o lo cruzan, recortándose para que no salgan de este. Y en el caso de polígonos, estos se recortan para restringir su extensión a la del polígono de recorte.

Un aspecto muy importante en el recorte de capas vectoriales es el tratamiento de las tablas de atributos asociadas a cada elemento. Salvo en el caso de puntos, en el caso de recortar alguna capa de otro tipo, las geometrías de esta se modifican, y la capa resultante contiene la geometría modificada y el mismo registro original asociado a ella.

Cuando la capa recortada es de puntos, no existe problema en esto, ya que no se da modificación alguna de las geometrías. El recorte es en realidad una operación de consulta y selección. La información de la tabla sigue correspondiéndose con la entidad geométrica, ya que ninguna de las dos ha cambiado. Cuando se recortan líneas o polígonos, sin embargo, la situación es algo más compleja.

Supongamos que la capa recortada es de polígonos tales como unidades administrativas. Si el registro original contiene información tal como por ejemplo el nombre de la unidad o su código postal asociado, no existe ningún problema, ya que estos valores se aplican de igual modo para la parte de cada polígono que quede tras el recorte. Una situación muy distinta la tenemos, sin embargo, cuando la tabla de atributos contiene valores tales como el área, el perímetro o el número de habitantes de dicha unidad administrativa. En este último caso, dichos valores guardan una relación directa con la geometría, y al cambiar esta deberían modificarse igualmente.

La operación de recorte no actúa sobre las tablas de atributos, ya que no posee información suficiente para poder hacerlo. Tanto el área como el perímetro deben recalcularse para que la información de la tabla de atributos sea coherente con el nuevo polígono al que se encuentra asociada. Como ya sabemos, medir estas propiedades de un polígono es sencillo a partir de sus coordenadas, y debe simplemente tenerse esa precaución. Se tiene que el recorte de la capa no es únicamente una operación geométrica, sino que, según sean los valores de la tabla de atributos, debe operarse también con ellos para completar dicha operación.

El caso del número de habitantes es algo distinto al del perímetro o el área, ya que no es una propiedad puramente geométrica. Puede calcularse el número de habitantes del polígono recortado aplicando una mera proporción entre las superficies original y recortada, pero el valor resultante solo será correcto si la variable (es decir, el número de habitantes) es constante en el espacio. Entramos aquí en el problema de la falacia

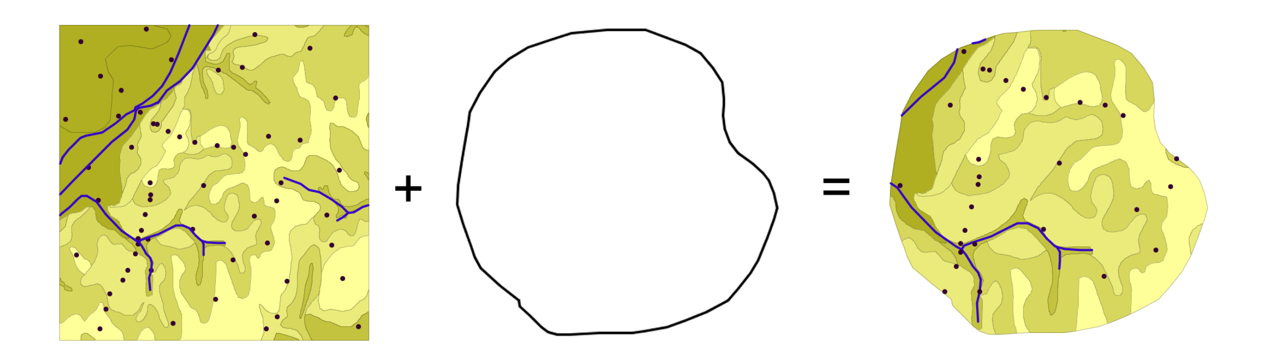


Figura 19.5: Recorte vectorial de un conjunto de capas de puntos, líneas y polígonos, con una capa de polígonos.

ecológica, el cual vimos en la sección 11.2.2, y que es de gran importancia a la hora de aplicar operaciones de solape.

Cuando la capa recortada es de líneas, debe considerarse de igual modo el hecho de que ciertas propiedades dependen directamente de la geometría, y por tanto definen la línea completa, no la parte de esta que queda tras el recorte. La longitud es un claro ejemplo de esto. Al no constituir regiones, el Problema de la Unidad de Área Modificable y otros problemas derivados no se han de considerar de la misma forma que en el caso de capas de polígonos, pero es necesario igualmente estudiar los valores de la tabla de atributos, para ver cómo el recorte de la capa afecta al significado de estos.

19.3.2. Diferencia

La diferencia es un operador contrario al recorte. En este último, como hemos visto, se mantienen en la capa resultante las geometrías de la capa recortada que entran dentro del área de recorte definida por otra capa adicional. En la diferencia, el proceso es semejante, pero en este caso las zonas que se mantienen son las que *no* entran dentro de la zona definida por la capa de recorte.

Puede entenderse como la realización de un recorte, pero en lugar de utilizando un conjunto de polígonos de recorte, empleando su complementario.

Mientras que el recorte era útil para restringir la información de una capa vectorial a un área dada, la diferencia es útil cuando deseamos excluir dicho área de la capa. Por ejemplo, dada una zona de influencia de un cauce, recogida esta en una capa vectorial de polígonos, puede interpretarse de cara a una planificación del terreno como una zona no apta para la edificación. A la hora de llevar a cabo un estudio relativo a dicha edificación, es interesante eliminar las zonas de influencia, ya que no van a tenerse en cuenta de ahí en adelante al no ser aptas para la actividad analizada.

Por su similar naturaleza, todas las consideraciones anteriormente hechas para el caso del recorte deben igualmente tenerse presentes al aplicar la operación diferencia.

19.3.3. Intersección

La intersección es equivalente a una operación booleana Y (AND), ya que la capa resultante mantiene solo aquellas zonas para las que se dispone de información en ambas capas de entrada. Es decir, aquellas zonas donde hay entidades en ambas capas. En particular, estas entidades han de ser de tipo polígono.

A diferencia del recorte, la información empleada para crear la tabla resultante no proviene únicamente de una capa (la capa recortada), sino de ambas capas de origen. Por ello, se producen modificaciones en las geometrías, que se dividen (se «trocean») según sea la intersección con las geometrías de la otra capa, y también en las tablas de atributos. Los atributos de cada una de las nuevas entidades son todos los asociados a las entidades que han dado lugar a dicha entidad intersección. Puesto que solo se mantienen en la capa resultante las entidades donde exista coincidencia, siempre habrá información en ellas sobre ambas capas. Es decir, la tabla de atributos resultante tiene tantos campos como el conjunto de las dos capas de partida.

Puede verse un sencillo ejemplo en la figura 19.6.

El hecho de que en la tabla aparezcan dos campos con el mismo nombre no implica que deban en modo alguno «mezclarse» los valores de estos. Simplemente existirán en la tabla resultante dos campos distintos con un nombre coincidente.

Al igual que en el caso anterior, de existir capas con valores que guarden relación con el área de cada polígono, los valores en las tablas pierden su significado al llevar a cabo la intersección. De hecho, podrían existir en ambas capas de origen sendos campos con un valor de área, que aparecerían ambos en la tabla de la capa resultante. Ninguno de ellos, no obstante, sería aplicable a la entidad a la que hacen referencia.

Este caso que acabamos de ver es idéntico en concepto al que veíamos en el apartado 15.2.1 sobre combinación de capas mediante el álgebra de mapas. Si se recuerda lo visto entonces, era necesario codificar de forma particular los valores en las capas de entrada para llevar a cabo correctamente la operación de combinación. Esto era así porque la capa resultante debía contener la información de las dos capas de entrada (tipo de suelo y uso de suelo, en aquel ejemplo), pero debía almacenar dicha información en un único valor, ya que una capa ráster tiene tan solo un valor asociado a cada celda.

Al trabajar con capas vectoriales, no existe esa limitación, y cada entidad puede llevar asociados tantos campos como se quiera. Por esta razón, no es necesario «preparar» las capas de entrada y modificar sus valores, ya que toda la información que contienen puede incorporarse a la capa resultante sin más que añadir los campos en los que se encuentra.

19.3.4. Unión

Si la intersección se puede asimilar a un operador lógico Y (AND), la unión es semejante al operador lógico O (OR). En la capa resultante del proceso aparecen todas las geometrías de la intersección y, junto a estas, también aquellas que corresponden a las zonas que aparecen únicamente en una de las capas de origen. Al cruzar estas capas, y al igual que en el caso de la intersección, sus geometrías se «trocean», pero en este caso todos esos «trozos» obtenidos aparecen en la capa resultante, y no solamente algunos de ellos.

De esta forma, y al unir dos capas de polígonos, encontraremos en la capa resultante zonas que están cubiertas por uno de ellos perteneciente a la primera capa, o bien por uno de la segunda capa, o bien por polígonos de ambas capas. Si,

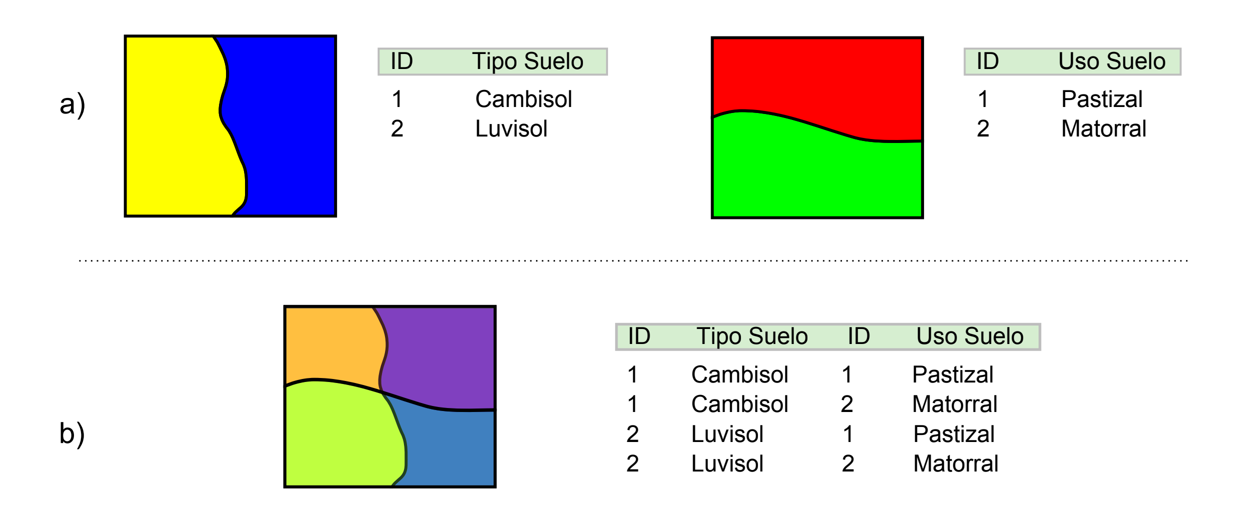


Figura 19.6: Interseccion entre dos capas de polígonos, mostrando las geometrías y la tabla de atributos resultante.

por ejemplo, esas capas representan zonas de influencia de dos procesos, podremos de este modo conocer cuáles afectan a cada una de las geometrías resultantes, y saber si, dentro del perímetro de dicha geometría, estamos dentro de la zona de influencia de ambos procesos, o bien solo en la de uno de ellos.

En general, la unión resulta de interés cuando estudiemos la existencia de dos fenómenos y queramos ver dónde tiene lugar al menos uno de dichos ellos. En la intersección buscábamos conocer en qué lugares tenían lugar los dos fenómenos simultáneamente.

La tabla de atributos correspondiente es igual a la de la intersección, con tantos campos como el conjunto de las dos capas de partida. En esta ocasión, no obstante, y por existir polígonos resultantes que no aparecerían en la intersección (zonas donde solo uno de los fenómenos representados se produce), aparecerán campos sin información, ya que no existen información suficiente para asignarse en esos casos.

La figura 19.7 muestra un ejemplo de unión de capas vectoriales.

19.3.5. Polígonos espúreos

Las operaciones geométricas de solape cruzan las geometrías de dos capas y calculan los elementos resultantes de la combinación de estas. Al llevar esto a cabo, es probable que en ciertos lugares existan líneas de estas capas que debieran coincidir pero que, debido a imprecisiones en su digitalización o a la precisión particular de cada capa, no lo hagan exactamente. Es decir, una misma realidad queda registrada de formas distintas en las diferentes capas de origen.

A la hora de efectuar la combinación anterior, esa falta de coincidencia va a dar lugar a polígonos adicionales que no deberían estar ahí. Estos, además, serán de pequeño tamaño, ya que las imprecisiones son pequeñas en relación al tamaño de las geometrías intersecadas. Estos polígonos son artificios fruto de las imprecisiones existentes

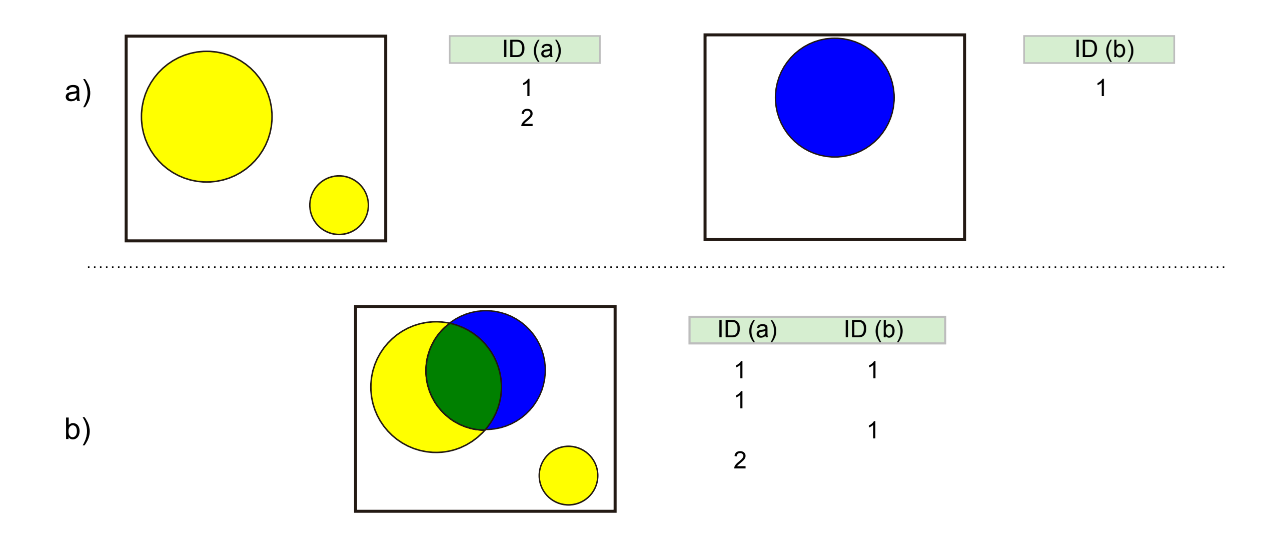


Figura 19.7: Unión de dos capas de polígonos, mostrando las geometrías y la tabla de atributos resultante.

en las capas sobre las que se realiza la operación geométrica en cuestión, y su eliminación es un paso complementario a dicha operación, el cual debe llevarse a cabo igualmente.

La aparición de polígonos espúreos (también frecuentemente citados en su denominación inglesa, *sliver polygons*) puede evitarse a la hora de realizar la intersección, incorporando en los algoritmos correspondientes una cierta tolerancia que permita que líneas distintas puedan tratarse como idénticas(como deberían ser en realidad), siempre que la diferencia entre ellas no supere dicha tolerancia.

Otra solución es la eliminación *a posteriori*, tratando de localizar los polígonos espúreos diferenciándolos de aquellos que sí representen regiones reales que deben conservarse. Algunas de las características que suelen presentar habitualmente y que pueden emplearse para esa distinción son:

- Pequeño tamaño
- Forma alargada
- Bajo número de lados. Son polígonos simples con pocos segmentos.

La figura 19.8 muestra un ejemplo de la aparición de este tipo de polígonos.

19.4. Juntar capas

Juntar capas no es una operación geométrica propiamente dicha, ya que ninguna de las geometrías de las capas de entrada se ve alterada en lo que a sus coordenadas respecta. Es, no obstante, una operación de combinación, ya que, al igual que las anteriores, genera una nueva capa de datos espaciales vectoriales a partir de dos capas de partida. En realidad puede aplicarse sobre un número n de capas, aunque por simplicidad suponemos que estas son solo dos.

El resultado es una nueva capa que contiene la información de las dos capas de entrada, es decir todas las entidades que se encuentran en una u otra de estas. Sobre dichas entidades no se realiza ningún análisis geométrico, y el hecho de que estas intersequen o no carece de relevancia

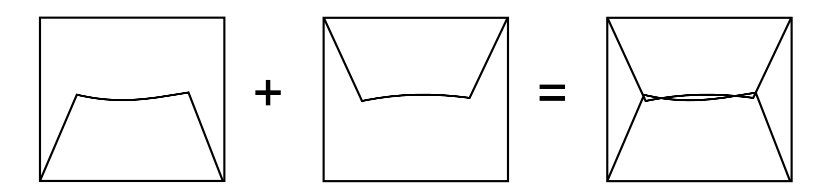


Figura 19.8: Aparición de polígonos espúreos de pequeño tamaño tras el solape entre capas con discrepancias.

para el resultado. Las relaciones espaciales entre entidades de ambas capas no se tienen en cuenta.

Por ello, si dos entidades por ejemplo poligonales, una de cada una de las capas de partida, se intersecan, ambas aparecerán en la capa resultante como tales, sin verse afectadas. En la zona de intersección habrá dos polígonos distintos. Esto no sucedía en las operaciones de solape vistas anteriormente.

La parte principal de la operación no es, por tanto, la relativa a las geometrías, ya que estas simplemente se «reúnen» en una sola capa. La consideración más importante es la que respecta a la información asociada a la capa resultante, que proviene de las dos capas de origen y define realmente el significado de esta.

La tabla de la capa resultante contiene tantos elementos como existan en el conjunto de capas de partida. Si estas contienen respectivamente n y m elementos, la capa resultante tendrá $n + m$ entidades. Para cada elemento se recogen tantos campos como campos diferentes aparezcan entre las dos tablas. Las entidades de una de las capas, si no tienen valores para los campos provenientes de la otra —por no aparecer este campo en ambas— no tendrán valor alguno. Algunos SIG dan la opción de seleccionar qué capa es la capa principal, cuyos campos se emplearan para la capa definitiva. La información de las otras capas que se recoja en campos no existentes en dicha capa principal se perderá en la capa resultante. Con independencia de la implementación, el concepto es similar en todos los casos.

Para comprender mejor esta operación, puede verse un ejemplo en la figura 19.9.

Para aplicar esta operación de forma coherente y que la tabla generada según lo anterior tenga pleno sentido, ambas capas de origen tienen que contener no solo el mismo tipo de entidades, sino también información de índole similar. Salvo en contadas ocasiones, no tiene sentido unir, por ejemplo, una capa de polígonos y otra de líneas, y será mejor mantenerlas independientes a todos los efectos. De modo similar, tampoco tiene sentido unir una capa de polígonos con valores de uso de suelo y otra con límites administrativos, ya que las tablas de datos de estas serán bien distintas y el resultado será poco coherente. La similitud en cuanto al tipo de entidad no garantiza que la operación tenga sentido.

La operación de juntar capas es útil en muchas circunstancias en las cuales se dispone de datos geográficos de distintas procedencias o que, por su propio origen, vienen divididos en partes. Un caso frecuente es el de disponer de la información por hojas coincidentes con la cartografía clásica, tales como las hojas proporcionadas por los Institutos Geográficos o instituciones similares. Si la zona estudiada cubre varias de estas hojas, tendremos los datos divididos en tantas capas distintas como hojas cubiertas. Sin embargo, lo ideal sería tenerlas todas en una única capa.

Esta conveniencia no solo es relativa al manejo de la capa, sino también para otros aspectos tales como el análisis o incluso la visualización. A la hora de analizar los datos, muchas formula-

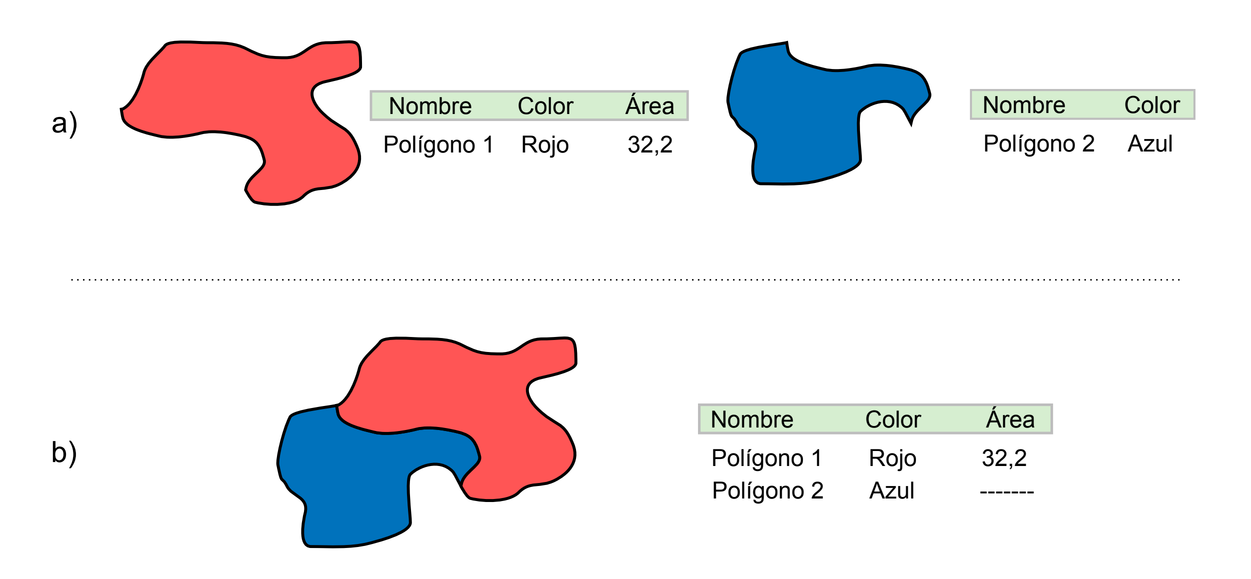


Figura 19.9: Ejemplo de las tablas y entidades resultantes tras juntar dos capas.

ciones utilizan no solo la geometría o los valores asociados a esta, sino también la relación con otros elementos de la misma capa. Tal es el caso en el análisis de redes, por ejemplo. Si la red de carreteras que recorre esas hojas en las que se sitúa la zona de estudia se encuentra en diversas capas, no se refleja la conectividad entre las carreteras de distintas hojas. Deben juntarse en una única antes de poder analizarlas conjuntamente.

Otras operaciones no requieren de esa conectividad, pero el aplicarlas sobre la información contenida en las capas implica hacerlo tantas veces como capas existan. En lugar de calcular, por ejemplo, un área de influencia de dichas capas, hay que calcular ese mismo área para cada capa. Después estas zonas de influencia podrían juntarse en una sola capa, pero resulta más lógico hacerlo a priori y después operar.

Juntando las capas, la visualización también se ve afectada. Además de no poder analizar conjuntamente ese conjunto de carreteras cuando están en capas separadas, tampoco pueden modificarse conjuntamente los atributos de representación. Cambiar el color o el grosor de las líneas que representan las carreteras implica cambiar ese color o grosor para cada una de las capas, para que el resultado sea visualmente homogéneo.

Si la paleta de colores no es de tipo absoluta y no utiliza una tabla de asignación, sino que adapta una rampa de colores entre los valores mínimos y máximos del campo utilizado para asignar los colores, es absolutamente necesario unir las capas que queramos para conseguir un resultado coherente. De otro modo, la misma rampa de colores no representará lo mismo en cada capa, ya que los máximos y mínimos entre los que se adapta serán distintos para cada parte (cada capa), y en el caso más habitual distintos asimismo de los correspondientes al total de los datos de la zona de estudio. Veremos un ejemplo de esto en el capítulo 23.

Pueden juntarse capas no solo porque abarquen áreas distintas con una misma información, sino también si cubren el mismo área pero con informaciones distintas. Estas informaciones deben compartir, no obstante, algún rasgo común.

Una capa de carreteras y otra con caminos pueden juntarse para formar una capa con las vías existentes en la zona de estudio.

19.5. Modificaciones basadas en atributos. Disolución

Las tablas de atributos pueden emplearse para definir la forma en que se realiza una operación geométrica. En lugar de ser elementos pasivos que no se ven modificados tras la operación, los atributos pueden ser quienes aporten la información necesaria para establecer la manera de modificar las entidades de entrada.

Una operación muy frecuente en este sentido es la llamada *disolución*. Esta operación recibe este nombre debido a que une polígonos con atributos comunes y «disuelve» las fronteras existentes entre ellos en una única entidad. No es necesario que exista una frontera entre los polígonos (es decir, que sean contiguos) ya que pueden almacenarse en una capa vectorial entidades compuestas por varios polígonos disjuntos. Tal es el caso, por ejemplo, de una entidad poligonal que represente a España, que contendrá no solo el polígono de la península, sino también los de las islas que pertenecen igualmente al país. Para todos ellos existe un único registro en la tabla de atributos asociada.

La aplicación de la operación disolver hace que todos aquellos polígonos que tengan asociado un determinado valor en uno de sus atributos pasen a constituir una nueva y única entidad, ya sea esta de un solo polígono o varios disjuntos.

Este es el caso que aplicábamos, por ejemplo, a la hora de simplificar las zonas de influencia. En dicho caso, se unen simplemente por contigüidad espacial todas las zonas generadas, asumiéndose por tanto que todas tienen algún valor común en sus atributos.

No obstante, pueden disolverse las entidades según distintos grupos, tantos como valores distintos de un atributo existan en la capa. Un ejemplo muy representativo en este sentido es obtener a partir de una capa con provincias una nueva con polígonos que representen comunidades autónomas. Es decir, agrupar un tipo de división en otra a una escala mayor. Para ello, por supuesto, debe existir información sobre a qué comunidad autónoma pertenece cada provincia, para poder aplicar la condición que permita seleccionar los polígonos a disolver.

En la figura 19.10 se muestra un ejemplo de lo anterior.

Al igual que en otras operaciones ya vistas, la tabla de atributos de la capa resultante merece atención aparte. En esta ocasión, existe un comportamiento diferente según si el atributo es numérico o no. Si el campo no es de tipo numérico, la unión de n entidades en una única implica «juntar» la información n valores no numéricos. Puesto que estos no permiten operaciones de tipo matemático, no es posible utilizar esos valores y obtener un valor para ese campo en la nueva capa. Por ello, estos campos no aparecerán en la tabla resultante.

En el caso de campos numéricos, pueden emplearse los datos de las capas de partida, aplicando operaciones diversas según sea la naturaleza de la variable. Por ejemplo, para un campo con el número de habitantes de cada término municipal, si aplicamos una operación de disolución y obtenemos una capa de comunidades autónomas, el valor de población de cada entidad resultante (cada comunidad autónoma), será la suma de los valores de los polígonos que han sido «disueltos» para obtener dicha entidad.

Si el campo en cuestión recoge la estatura media de la población, el valor resultante deberá ser una media ponderada de los valores de cada término, utilizando el número de habitantes como

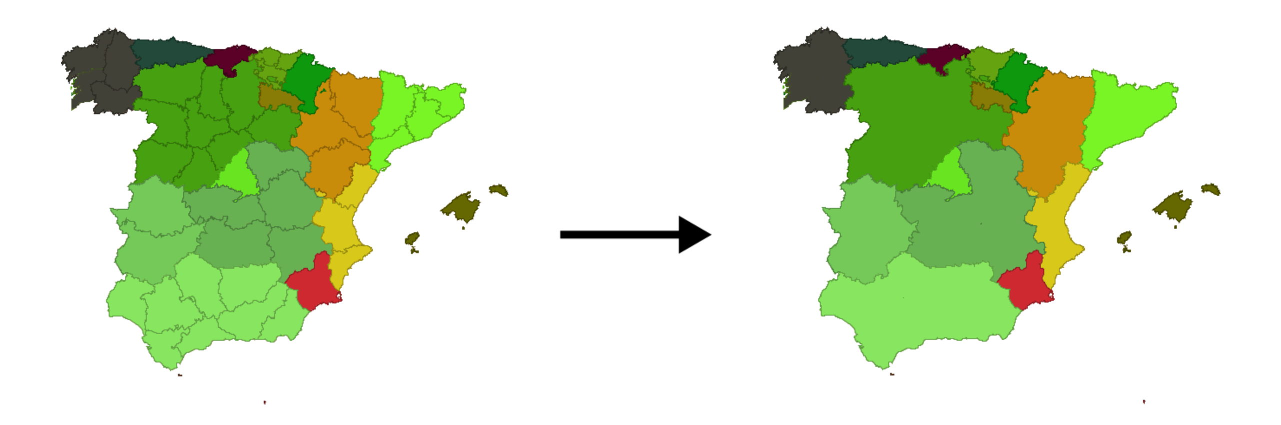


Figura 19.10: Obtención de una capa de comunidades autónomas a partir de una de provincias mediante un proceso de disolución. En la capa original se han representado los polígonos en función de la comunidad a la que pertenecen. Aunque aparecen con el mismo color, son polígonos independientes.

ponderación. Si en lugar de la media se recoge la estatura máxima, el máximo de todos los valores de los términos será el valor a incluir en la tabla de atributos de la capa resultante en ese campo.

Una vez más, es necesario considerar la naturaleza de la variable para establecer la forma de combinar los valores. Al hacerlo, y al igual que en otros casos, no deben perderse de vista los efectos derivados de la agregación que llevamos a cabo, los cuales ya conocemos.

19.6. Contornos mínimos

Dado un conjunto de puntos, una de las operaciones geométricas más comunes que pueden llevarse a cabo es la delimitación de un contorno mínimo que los englobe. Conocer el espacio ocupado por este contorno puede ser útil para evaluar diversos parámetros tales como la zona de cobertura del fenómeno representado por dichos puntos.

Existen diversas formas de contornos mínimos, entre las que cabe destacar.

- Envolvente convexa mínima
- Rectángulo mínimo
- Círculo mínimo

19.6.1. Envolvente convexa mínima (convex hull)

La envolvente convexa mínima (habitualmente citada como *convex hull*, su denominación en inglés) es la más común de las envolventes. Define el polígono convexo de menor área dentro del cual se contienen todos los puntos del conjunto, y su significado tanto geográfico como geométrico es de gran importancia.

Resulta fácil visualizar el concepto de esta envolvente si suponemos que rodeamos los puntos con una banda elástica. La forma que define dicha banda es la envolvente mínima convexa. Como puede verse en la figura 19.11, es sencillo trazar este polígono visualmente a partir de los puntos, pero su cálculo numérico es sumamente más complejo de implementar de lo que en apariencia puede pensarse. Los detalles acerca de algoritmos para el cálculo de esta envolvente pueden encontrarse, por ejemplo, en[271].

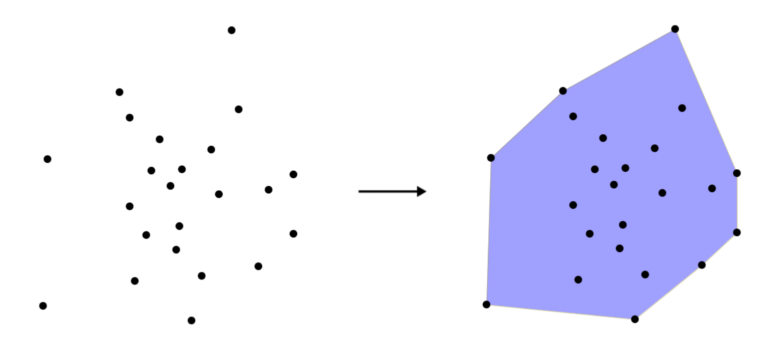


Figura 19.11: Contorno convexo mínimo.

La envolvente convexa delimita el área dentro de la cual se puede inferir el comportamiento de una variable a partir de una serie de muestras. Por ejemplo, en el caso de interpolar un conjunto de valores tal y como vimos en el capítulo 14, los valores estimados dentro de la envolvente convexa mínima son producto de una interpolación, mientras que estimar valores fuera de dicha envolvente constituye en realidad un proceso de extrapolación.

El polígono que define a la envolvente mínima convexa puede ser empleado asimismo como dato de entrada para otras operaciones geométricas. Dados, por ejemplo, una serie de puntos en los que ha aparecido una enfermedad infecciosa,puede calcularse la envolvente mínima convexa y a partir de ella calcular una zona de influencia con una distancia definida en función de la capacidad de propagación de la enfermedad. El nuevo polígono resultante representa la región que puede verse afectada por dicha enfermedad.

Una serie de envolventes convexas sucesivas sobre un conjunto de puntos forma una teselación en capas en forma de cebolla[1], de utilidad para la realización de diversos análisis estadísticos sobre dichos puntos.

[1] *Onion peeling*, en inglés

19.6.2. Círculo mínimo

El círculo mínimo es aquel circulo que contiene a todos los puntos de un conjunto dado con el menor radio posible.

El significado de este círculo mínimo es variado, ya que tanto el centro como el radio del mismo se prestan a diversas interpretaciones. Por ejemplo, si suponemos un terreno plano y una serie de núcleos de población (cada uno de ellos representado por un punto), y queremos situar una antena para dar cobertura a la región definida por esos puntos, el centro del círculo mínimo es una buena opción. Esto es así porque minimiza la distancia a la que se sitúa el punto más alejado, y por tanto minimizará la fuerza de la señal necesaria para ofrecer esa cobertura completa. La intensidad de la señal se puede calcular en función del radio del círculo.

Un análisis similar es frecuente en el terreno militar. Para un conjunto de n puntos a atacar, el emplazamiento de una bomba en el centro del circulo mínimo permitirá que ese ataque afecte a todos los puntos con una cantidad mínima de explosivo. Dicha cantidad está en relación, al igual que en el caso anterior, con el radio del círculo.

La construcción de algún elemento de servicio compartido, tal como un colegio o un hospital también puede analizarse empleando el círculo mínimo. Si situamos dicho elemento en el centro, garantizamos que la distancia del usuario más lejano de dicho servicio es mínima. Cualquier otro emplazamiento implicaría que existe un usuario en peores condiciones, que tendrá que recorrer una distancia mayor para llegar a ese colegio u hospital.

Como vemos, el círculo mínimo es una herramienta útil para la localización y emplazamiento de distintos elementos. Estos problemas, no obstante, son más complejos en general, ya que implican la localización de varios elementos, o bien existen otros previos que han de considerarse, así como diversos factores externos.

Además de esta interpretación geográfica, el círculo mínimo aporta otro tipo de información. Es habitual, por ejemplo, que los puntos que se sitúan en el borde del mismo sean descartados, ya que constituyen los más alejados y en cierta medida son elementos extremos poco representativos del conjunto.

Existen muchos algoritmos para el cálculo del círculo mínimo, que escapan al ámbito de este texto. Para saber más al respecto, puede consultarse[272] o [273].

Es de reseñar que los punto situados en el borde del círculo mínimo siempre pertenecen a su vez a la envolvente mínima convexa. Por ello, el problema del cálculo del circulo mínimo para un conjunto de n puntos puede reducirse al cálculo de dicho círculo para el subconjunto de puntos que componen la envolvente mínima convexa.

19.6.3. Rectángulo mínimo

El rectángulo mínimo es el rectángulo de menor área que cubre todos los puntos de un conjunto.

Pare el cálculo del rectángulo mínimo se aplica el hecho de que al menos un lado de este se sitúa sobre un lado de la envolvente mínima convexa. Por ello, basta comprobar los distintos rectángulos que pueden construirse sobre dicha envolvente, y tomar el de menor área.

Generalmente, el rectángulo de menor área coincide con el de menor perímetro, pero no siempre es así. De cualquier modo, este último también cumple la condición citada con respecto a la envolvente convexa, por lo que su cálculo puede hacerse por un procedimiento idéntico.

19.7. Generalización de líneas

Como ya sabemos, toda la información vectorial la almacenamos en ultima instancia como un conjunto de puntos, ya sean aislados o unidos mediante segmentos o curvas para conformar líneas o contornos de polígonos. Una transformación habitual en el caso de líneas o polígonos consiste en la modificación de ese conjunto de puntos de tal modo que se reduzca su número pero se preserve en la medida de lo posible la información que originalmente contenían.

Este proceso de simplificación es parte de la *generalización* de líneas, y es importante dentro de los Sistemas de Información Geográfica, tanto para la representación de datos como para su análisis o su simple almacenamiento, como ya vimos al inicio de este libro.

Las razones por las cuales puede resultar de interés llevar a cabo un proceso de generalización de líneas son diversas, y entre ellas cabe destacar las dos siguientes [23].

- **Reducción del tamaño de los datos**. Una reducción del número de puntos elimina puntos en muchos casos superfluos, de tal forma que la capa simplificada presenta la misma utilidad pero ocupa un espacio menor.

- **Reducción del tiempo de proceso**. La capa generalizada se maneja de forma más rápida en operaciones tales como la representación en pantalla, la impresión, o la realización de otros cálculos. En términos generales, todos los cálculos con la línea generalizada, como por ejemplo el trazado de una zona de influencia o de cualquier otro de los procesos vistos en este capítulo, se efectúan con un menor costo de proceso, ya que requieren el análisis de un menor número de puntos. Otros procesos tales como la conversión de esa capa en una capa ráster también experimentan una ganancia en rendimiento.

En ocasiones, la simplificación puede implicar la reducción de elementos más allá de puntos aislados, tal y como operan los algoritmos que a continuación veremos. Eliminando puntos a lo largo de una línea puede lograrse el resultado buscado, reduciendo el detalle longitudinalmente, pero un cambio de escala puede también implicar la necesidad de eliminar no únicamente puntos, sino líneas completas. Por ejemplo, si una capa de líneas recoge con detalle una vía mediante dos líneas, una para cada borde de la misma, no tiene sentido emplear una capa de tal detalle para un mapa a una escala tal como, por ejemplo, 1:200000. En este caso, puede sustituirse el par de líneas anteriores por una única, ya que la variación en el resultado no será perceptible. La simplificación de las líneas en este caso debe operar sustituyendo dos líneas por una única.

Si esa carretera queda recogida mediante un polígono, puede simplificarse mediante un proceso de «adelgazamiento» que convierta este en una línea central. La obtención de este eje del polígono se hace con un proceso que es similar a la zona de influencia hacia el interior de un polígono, la cual veíamos al principio de este mismo capítulo.

Una operación también relacionada con la generalización de líneas es el *suavizado*. En muchas ocasiones, las líneas de una capa vectorial son excesivamente angulosas y no presentan un aspecto natural, muy distinto del trazo suave con que un cartógrafo trazaría estas al realizar un mapa. Alterando el conjunto de puntos de la línea de forma similar a lo visto anteriormente, puede lograrse un redondeo en las curvas definidas por este. Dicho redondeo puede buscarse con meros fines estéticos, pero también para eliminar variaciones bruscas o desviaciones locales, manteniendo tan solo la forma general de la línea. Este es, por ejemplo, el proceso que debe realizarse si deseamos utilizar las líneas a una escala de menor detalle que la que originalmente se empleó en su creación.

El suavizado de líneas es también útil como preparación de datos de líneas procedentes de un proceso de digitalización. Aunque muchos elementos naturales tienen formas redondeadas, es habitual que la digitalización genere elementos más angulosos que el objeto real que se digitaliza.

Algo similar sucede cuando las operaciones de vectorización se realizan de forma automática, tales como las que vimos en en capítulo 18. El suavizado de las líneas mejora en tal caso la calidad de estas, no solo en su aspecto estético, sino también en su similitud con el objeto modelizado, ya que se trata de formas más naturales(Figura 19.12).

19.7.1. Métodos

Los algoritmos para la generalización de líneas son muy diversos y tienen cada uno sus propias características de precisión y rendimiento. La forma más simple de generalización consiste sencillamente en eliminar puntos sin considerar la relevancia de estos dentro de la línea. Esta eliminación puede realizarse de forma sistemática (eliminar un punto de cada n), o bien aleatoria. La

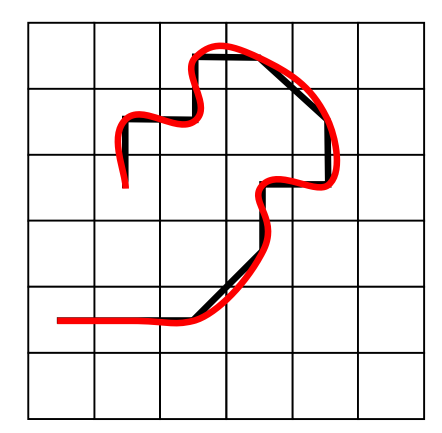


Figura 19.12: Suavizado de una línea procedente de vectorización a partir de una capa ráster. En negro, línea original con esquinas angulosas. En rojo, línea suavizada.

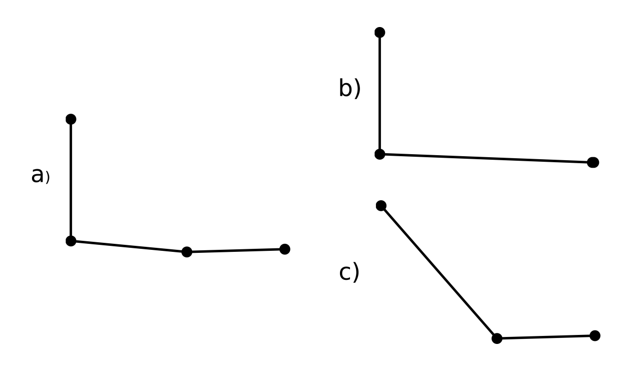


Figura 19.13: La eliminación de puntos de la línea original (a) puede dar lugar a líneas muy similares (b) o muy distintas (c), ya que no todos los puntos tienen la misma importancia.

magnitud del proceso de generalización se mide por el número total de puntos eliminados.

No obstante, no todos los puntos de un línea tienen la misma importancia y aportan la misma cantidad de información. Algunos puntos pueden resultar redundantes, mientras que otros pueden ser cruciales para la forma del trazado. Como puede verse en la figura 19.13, eliminar algunos puntos puede no tener apenas efectos sobre la línea original o bien resultar en una variación drástica de su forma. Un mismo número de puntos eliminados puede dar lugar a líneas muy similares o muy distintas a la original, según sea el caso, por lo que esta medida de la simplificación no es una medida de cuan fidedigna es la línea resultante.

Puesto que no todos los puntos deben ser considerados de igual modo, existen métodos de simplificación que consideran la naturaleza propia del punto dentro de la línea y estudian esta como entidad en su conjunto. Esto garantiza resultados mejores que con la mera eliminación sistemática de puntos intermedios.

Podemos clasificar las rutinas de generalización de líneas en los siguientes bloques [275].

- **Rutinas de vecindad inmediata**. Analizan cada punto y los inmediatamente anteriores y posteriores, estudiando los ángulos formados o las distancias.
- **Rutinas de vecindad acotada**. Estudian una serie de puntos vecinos, no necesariamente restringida a los inmediatos pero sí con algún umbral máximo de alejamiento o número de puntos.
- **Rutinas de vecindad no acotada**. Estudian los puntos vecinos y la región analizada depende de diversos factores, aunque no existe una dimensión máxima.
- **Rutinas globales**. Analizan la línea de forma global.
- **Rutinas que preservan la topología**. Además de las propiedades geométricas, ya sea a nivel global o local, estudian la conectividad y topología del conjunto de líneas para garantizar que la línea simplificada preserva dichas propiedades.

Uno de los algoritmos más habitualmente utilizados es el propuesto por[274], que pertenece

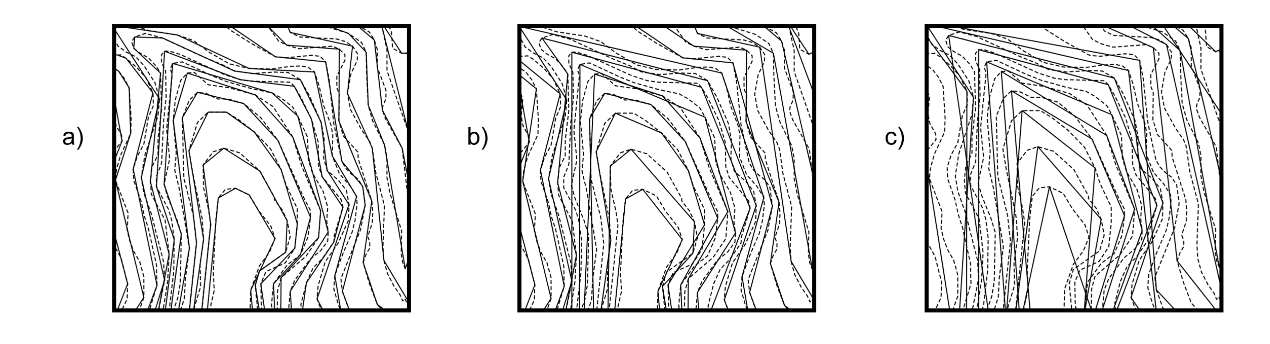


Figura 19.14: Generalización de un conjunto de líneas (en trazo punteado) mediante el algoritmo de [274] para valores de tolerancia de 10 (a), 20 (b) y 50 metros (c).

al grupo de los globales. La intensidad del proceso de generalización se establece mediante un valor de tolerancia que indica la máxima desviación que se permite entre la línea original y la simplificada.

En la figura 19.14 puede verse cómo una línea original se simplifica de diversas formas al aplicar tolerancias crecientes con este algoritmo.

Respecto a los algoritmos de suavizado, estos pueden dividirse en tres grupos principales [275]:

- **Media entre puntos**. Se consideran un número n de puntos vecinos, y en base a ellos se calcula la nueva posición.
- **Ajuste de funciones matemáticas**. Ajustando a los puntos funciones que tengan un aspecto «suave», tales como *splines* o curvas Bézier.
- **Tolerancias**. Se establece una tolerancia y un umbral de precisión, y se ignoran los detalles a lo largo de la línea que salen de ese umbral.

En [276] pueden encontrarse detallados métodos de todas las familias anteriores.

19.8. Resumen

Las operaciones geométricas sobre entidades vectoriales constituyen en cierta forma el equivalente del álgebra de mapas ráster sobre las capas vectoriales.

Las operaciones más importantes son el cálculo de zonas de influencia y las denominadas operaciones de solape, que permiten combinar capas de diversas maneras. Entre estas encontramos las operaciones de intersección, unión, diferencia y recorte, así como el juntado de capas. Las tablas de valores asociadas a las capas resultantes de estos procesos deben considerarse y tratarse con precaución, ya que las transformaciones que se realizan sobre las geometrías pueden afectar a la validez de los datos que contienen.

Las tablas pueden también usarse para definir otro tipo de operaciones tales como la disolución, en la que se unen regiones que comparten algún atributo común.

Por último, las operaciones de generalización de líneas son útiles para disminuir el tamaño de los datos vectoriales, con las ventajas de manejo y proceso que ello conlleva.

Capítulo 20

Costes, distancias y áreas de influencia

20.1. Introducción

Desplazarse entre dos puntos cualesquiera del espacio implica un coste. Cuanto más alejados se encuentren estos, más tiempo llevará efectuar ese desplazamiento y se consumirá más energía, entre otras cosas. La distancia como tal, así como el tiempo o la energía gastada, son ejemplos de variables de coste.

Un coste expresa la resistencia ofrecida por el medio para desplazarse a través de él en un punto concreto.

El análisis de capas que contengan variables de este tipo nos permite calcular los costes globales de desplazarse a lo largo de toda la extensión de la capa, así como evaluar la mejor forma de hacerlo para minimizar dicho coste.

Las variables que podemos utilizar como coste son muy diversas, y este tipo de análisis es de gran utilidad para todos aquellos procesos que impliquen algún tipo de movimiento.

De este modo, el concepto de distancia puede «ampliarse», y con él todas sus implicaciones. Por ejemplo, las zonas de influencia no solo pueden definirse con la mera distancia euclídea, sino en base a otros factores. Así, podemos incorporar el hecho de que, mediando la misma distancia entre una celda y un conjunto de otras dadas, la facilidad de recorrer esta distancia sea diferente, y por tanto no ejerzan todas la misma influencia sobre la primera.

La delimitación de zonas de influencia sobre una base ráster permite no solo introducir estas ideas, sino también aportar más flexibilidad al proceso, pudiendo emplearse parámetros adicionales que condicionen la forma de llevarlo a cabo en cada punto.

20.2. Superficies de fricción

Una capa que contiene una variable de coste se conoce como *superficie de fricción*. Las variables de coste son de tipo cuantitativo y generalmente, aunque no siempre, continuas.

El valor de cada celda de una superficie de fricción indica el coste que supone recorrer dicha celda [277]. Puesto que la celda puede recorrerse en diversas direcciones, se establece que este coste se refiere a hacerlo en la dirección vertical u horizontal, no diagonal. Así, una capa que almacene la variable de coste «distancia» será una capa constante que contendrá en todas las celdas el tamaño de celda Δs.

No obstante, los costes son, por regla general, variables. Esto es, distintos en cada celda. Veamos algunos casos.

Una capa de pendientes, por ejemplo, es una capa de coste válida, ya que la dificultad de desplazarse sobre el terreno aumenta conforme lo hace la pendiente. De igual modo, el tiempo empleado en atravesar una celda, que será función de diversos factores, también representa un coste.

Una superficie de fricción puede contener también un valor que no exprese directamente un coste, pero esté relacionado con él y permita obtenerlo. Por ejemplo, la velocidad media de tránsito a través de una celda. Con este valor y el tamaño de celda se puede obtener el tiempo de tránsito.

El coste no es necesariamente una variable física de tipo continuo. Por ejemplo, podemos clasificar las distintas zonas de una capa en función de su atractivo visual. Invirtiendo estos valores obtenemos una variable de coste, que nos indica la «dificultad» de atravesar la celda en función de su atractivo. Las zonas con menor belleza suponen un mayor coste. Esto nos puede servir para calcular rutas agradables, ya que las rutas óptimas en este caso son aquellas que pasan por un menor numero de zonas estéticamente no agradables.

Este tipo de variables hacen referencia a los denominados espacios *subjetivos* [278], en contraposición a los espacios absolutos sobre los que se registran las medidas de distancia o tiempo que constituyen las fricciones empleadas más frecuentemente.

Aunque la capa con la superficie de fricción cubre toda una extensión dada, es probable que puntos de esta extensión no sean transitables. Si suponemos que el coste viene expresado por la pendiente, hay zonas tales como embalses donde, con independencia de su pendiente, no podemos cruzar. Una solución habitual es asignar a las celdas en estas zonas un valor muy elevado (coste infinito), que haga que no sea viable transitarlas. Una opción más correcta es asignar valor de sin datos a estas zonas, de forma que queden directamente excluidas del cálculo. Esto puede emplearse para excluir en el posterior trazado de rutas óptimas aquellos puntos por los que no se desee transitar, bien sea por los propios valores de la variable de coste o bien por otras razones.

Un ejercicio sencillo para comenzar a trabajar con superficies de fricción es el mostrado en la figura 20.1. Sobre la pequeña capa ráster de la figura con valores de fricción, se representa una ruta entre dos de sus celdas. Podemos calcular el coste total de recorrer la ruta sumando los costes de cada movimiento entre celdas consecutivas. Para los desplazamientos en dirección horizontal o vertical, el desplazamiento tiene un coste dado por la expresión:

$$c = \frac{c_1 + c_2}{2} \tag{20.2.1}$$

siendo c_1 y c_2 las fricciones respectivas de las celdas entre las cuales se produce el desplazamiento.

Si nos movemos en dirección diagonal, el coste que este movimiento supone es

$$c = \sqrt{2}\frac{c_1 + c_2}{2} \tag{20.2.2}$$

Con lo anterior, el coste de la ruta señalada en la figura 20.1, comenzando en la celda de la esquina superior izquierda, es igual a

$$\frac{1}{2} + \frac{3}{2} + \frac{3\sqrt{2}}{2} + \frac{2\sqrt{2}}{2} + \frac{2}{2} + \frac{2}{2} + \frac{2}{2} + \frac{1}{2}$$

20.3. Superficies de coste acumulado

Aunque de interés, el análisis anterior es muy simple, y no revela todo el potencial de la superficie de fricción. Mediante un análisis más intenso, podemos, por ejemplo, y dados los mismos puntos, calcular la ruta que haga mínimo el coste total de desplazamiento.

De igual modo, el análisis de la superficie de fricción es la base para resolver otros problemas de rutas óptimas. Dada una serie de puntos posibles de destino y un punto de inicio, podemos calcular cuál es el punto más cercano (cercanía

1	3	5	9	6	0
3	5	2	2	3	4
7	2	8	1	1	4
2	5	6	6	5	3
1	9	3	4	2	8
5	7	2	1	8	6

Figura 20.1: Una superficie de fricción permite calcular el coste de una ruta definida sobre ella.

en términos de coste mínimo, no de distancia euclídea), y calcular la ruta óptima y el coste de esta.

Para ello, debemos convertir la capa con los costes unitarios por celda en una *capa de coste acumulado*. El coste acumulado de cada celda representa el valor del coste total a recorrer desde dicha celda a la celda de destino más cercana, desplazándose por la ruta óptima. Para una superficie de fricción con distancias euclídeas, y un conjunto de n puntos, la capa de coste acumulado correspondiente indica la distancia al punto más cercano. Si utilizamos una superficie de fricción con tiempos de tránsito, la capa de coste acumulado indica el tiempo mínimo que se tardaría en alcanzar uno de esos puntos de destino (el más cercano en términos de tiempo).

De forma habitual, los puntos de destino se recogen en otra capa ráster, codificados con un valor concreto, o bien simplemente con valores cualesquiera y utilizando valores de sin datos en las celdas que no representan puntos de destino.

Para convertir una superficie de fricción en una superficie de coste acumulado en base a unos puntos de destino dados, se sigue el siguiente procedimiento [174, 279] :

- Para cada una de las celdas de destino, analizamos el coste de desplazarse a las adyacentes.
- Si estas no han sido analizadas aún o ya tienen un coste asignado pero es mayor que el calculado en el paso anterior, se les asigna el valor de dicho coste calculado.
- Desde este punto, repetimos el proceso pero en lugar de utilizar las celdas de destino, utilizando todas las celdas que han sido modificadas en la iteración anterior.
- El proceso se detiene cuando no se modifican nuevas celdas, ya que en este punto todas tienen asignado su coste acumulado mínimo.

Para visualizar gráficamente esta metodología, puede observarse que el proceso de ir seleccionando las celdas de análisis de cada iteración se asemeja a la propagación de las ondas al arrojar una piedra en un estanque, a partir de las celdas de destino iniciales que serían como el punto en el que cae la piedra. Contrariamente a lo que puede pensarse, a mayor numero de puntos de destino, menor tiempo de ejecución del algoritmo, de la misma forma que a mayor numero de piedras lanzadas sobre el estanque, menor tiempo tarda toda la superficie del mismo en ser alcanzada por alguna perturbación.

En la figura 20.2 podemos ver la capa de coste acumulado resultante de utilizar tres puntos de destino y una superficie de fricción constante. Puesto que la superficie de fricción es constan-

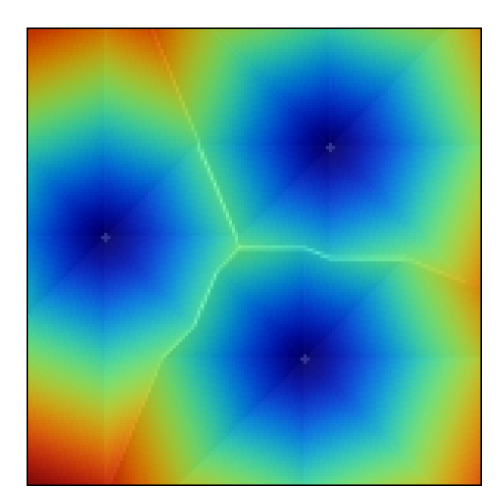


Figura 20.2: Capa de coste acumulado para tres puntos de destino y con una superficie de fricción constante. Para lograr una visualización más explícita se ha añadido un sombreado en función del propio coste.

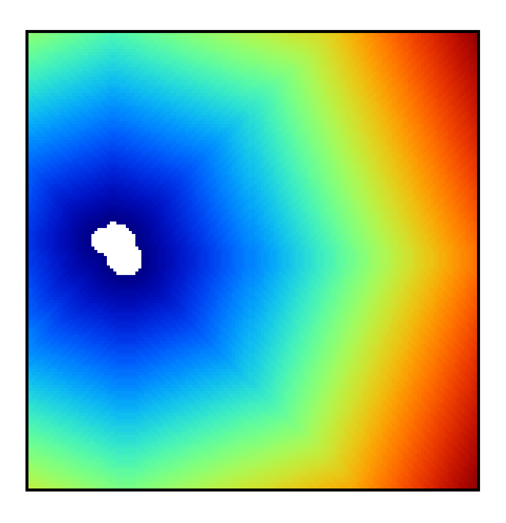


Figura 20.3: Superficie de coste acumulado representando el coste mínimo a un área de destino. Las celdas en blanco indican las celdas de destino, para las cuales el coste acumulado es nulo.

te, dicho coste acumulado es proporcional a la distancia.

Para calcular la superficie de coste acumulado necesitamos puntos de destino, codificados según algún criterio preestablecido, como ya se ha dicho. Puesto que trabajamos sobre una capa ráster, en realidad debemos definir celdas de destino. Estas, no obstante, no han de representar necesariamente localizaciones puntuales aisladas. Podemos establecer grupos de celdas de destino contiguas, que en realidad representan áreas de destino20.3. Esto nos permite recoger entidades lineales (por ejemplo, para calculo de costes de desplazamiento a una carretera) o de área, además de, por supuesto, elementos puntuales .

Uno de los problemas principales del cálculo con el método anterior es el debido a la limitación de los ángulos de movimiento. Al igual que veíamos para el modelo D8 de flujo, el hecho de que los movimientos se analicen en la ventana 3×3 obliga a que la dirección sea una de las definidas por las ocho celdas circundantes, es decir, siempre un múltiplo de 45°. Observando la figura 20.3, no es difícil percibir el efecto de esta limitación, del mismo modo que era sencillo detectar visualmente las deficiencias del modelo D8. Este efecto es especialmente patente al emplear una superficie de coste constante, tal y como se ha hecho en la figura anterior.

Para solventar en parte este inconveniente, una opción es analizar no solo esas ocho celdas, sino también algunas de la ventana 5×5 centrada en la misma celda. En particular, aquellas situadas a salto de caballo desde la celda central, por lo que esta conectividad se conoce como *vecindad de caballo*, en contraposición a la *vecindad de reina* que define el conjunto de celdas contiguas en la ventana 3×3 [280]. La figura 20.4 muestra esquemas de ambos tipos de vecindad.

20.3.1. Coste isotrópico *vs* coste anisotrópico

Al introducir el análisis de las superficies de fricción señalábamos la pendiente como una posible variable de coste. Los costes de un desplaza-

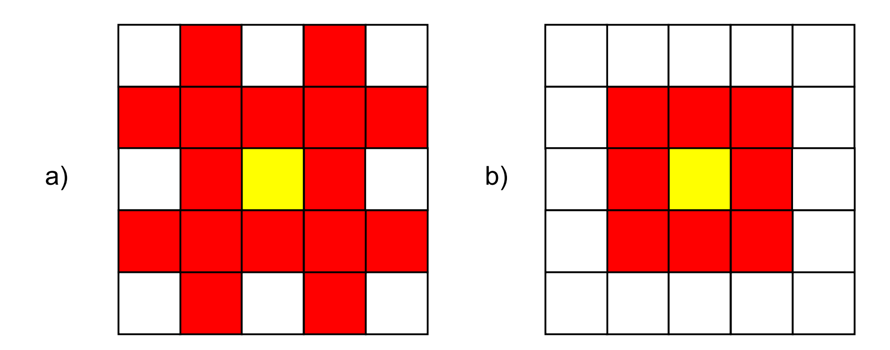


Figura 20.4: Tipos de vecindad. a) de caballo, b) de reina.

miento dependían de ese valor de la pendiente, así como de la dirección de desplazamiento, para aplicar el factor de corrección $\sqrt{2}s$ en caso de hacerlo diagonalmente. No obstante, además de considerar la dirección de desplazamiento para aplicar esa corrección en función de la distancia entre centros de celda, para modelizar el proceso de forma realista debemos considerarlo junto a la propia naturaleza de la variable de coste. En el caso de la pendiente, resulta obvio que no es igual recorrer la celda ascendiendo por la línea de máxima pendiente que hacerlo por la dirección de la curva de nivel. Es decir, el valor de coste en las celdas de una superficie de fricción no es único, sino que depende de la dirección.

Frente al coste isotrópico (igual en todas las direcciones) que hemos visto hasta el momento, más fácil de modelizar y comprender, encontramos fenómenos en los que la resistencia del medio no es la misma en todas direcciones. Este coste es de tipo anisotrópico. Junto al caso de la pendiente, otro ejemplo claro de esto es, por ejemplo, la resistencia que el viento ofrece al movimiento. Es obvio que el coste de desplazamiento no será el mismo si vamos en contra del viento que si vamos a favor. En general, la mayoría de procesos son de tipo anisotrópico, aunque muchos de ellos pueden simplificarse y estudiarse como procesos isotrópicos.

El concepto de superficie de fricción debe extenderse para acomodar este nuevo tipo de modelos. La capa con valores unitarios no es suficiente, ya que estos reflejan únicamente el coste en una dirección de todas las posibles. La solución habitual es recoger en la superficie de fricción los valores de coste máximo (el que se tendría desplazándose en contra del viento o hacia arriba de la ladera por la línea de máxima pendiente), y acompañar esta de una capa adicional con las direcciones en las que dicho coste máximo se produce. En el caso de la pendiente, esta dirección la podríamos obtener con la orientación en la celda, y en el caso del viento tomando la dirección opuesta a aquella en la que este sopla en cada celda.

Junto a esto, se define una función que, en base a la desviación angular respecto a la dirección de máximo coste, permite calcular los costes unitarios (fricción efectiva) en cualquier dirección. Con estos elementos ya podemos trabajar sobre la capa, estudiar el coste de una ruta concreta entre dos puntos, o crear una capa de coste acumulado de la misma manera que antes lo hacíamos con una capa de coste unitario isotrópico.

Una posible forma para la función anterior es la siguiente:

$$c = \frac{c_1 \cos^k(\alpha_1) + c_2 \cos^k(\alpha_2)}{2} \quad (20.3.1)$$

donde α_1 y α_2 son las diferencias entre el ángulo en que se produce el movimiento y los ángulos de máximo coste de las celdas, y k una constante.

Pueden adaptarse formulaciones más específicas si se conoce cómo modelizar un tipo de movimiento dado. Por ejemplo, la velocidad de propagación del fuego, conocida la velocidad en la dirección de máxima propagación, puede calcularse para las restantes según ciertas expresiones derivadas del análisis y modelización del fuego.

Dados los focos de un incendio, puede calcularse así el tiempo que tardarán en quemar toda el área definida por la superficie de fricción. En este caso, debe considerarse que los puntos no son de destino, sino de partida, con lo que la dirección a considerar es la opuesta. El trayecto cuyo coste se representa en la capa de coste acumulado no es para dicha celda *hasta* el foco más cercano, sino *desde* este.

El caso presentado de la pendiente como factor que condiciona el desplazamiento a pie puede expresarse con más precisión según la siguiente ecuación [281]:

$$T = a\Delta S + b\Delta H_1 + c\Delta H_2 + d\Delta H_3 \quad (20.3.2)$$

siendo T el tiempo empleado en el recorrido, ΔS la distancia recorrida, ΔH_1 la distancia vertical recorrida cuesta arriba, ΔH_2 la distancia vertical recorrida cuesta abajo con pendiente moderada y ΔH_3 la distancia vertical recorrida cuesta abajo en pendiente pronunciada. La pendiente moderada va de 5°a 12°y la pronunciada es aquella con ángulo mayor de 12°. Los valores propuestos de las constantes son [281] $a = 0{,}72, b = 6{,}0, c = 1{,}9998, d = -1{,}9998$.

Calculando la pendiente existente entre dos celdas entre las que se analice el coste, podemos así estimar el tiempo empleado.

En ocasiones, el movimiento no es posible en todas las direcciones. Es decir, en ciertas direcciones el coste es infinito. Por ejemplo, en el caso de modelizar el movimiento de un flujo aplicando un algoritmo como el D8 como veíamos en 16.6.1. De las ocho celdas hacia las que puede darse el movimiento, solo una de ellas es posible. Las restantes tendrían un coste infinito, pues el movimiento en esa dirección es inviable.

Aplicando este concepto es posible calcular una capa de distancias a un cauce, pero que esta distancia no sea euclídea, sino la seguida por el flujo desde cada punto hasta el punto en el que se unen con dicho cauce. Las celdas de cauce son en este caso las celdas de destino.

La figura 20.5 muestra un mapa de distancia euclídea a un cauce, así como otro de distancia hidrológica, pudiendo apreciarse la diferencia entre ambas.

Un problema adicional de trabajar con superficies de fricción anisotrópicas es la combinación de varias fuerzas resistentes.

Si las superficies de fricción son isotrópicas, y son varios los factores que dificultan el desplazamiento, combinar estos es tan sencillo como sumar las capas correspondientes. Si la influencia no es equivalente, pueden normalizarse o bien ponderarse, pero la capa resultante se obtiene con una mera suma y sigue siendo una superficie de fricción isotrópica.

Consideremos ahora el caso de dos variables de coste anisotrópico tales como el viento y la pendiente. En este supuesto no podemos sumarlas, ya que es necesario considerar también las direcciones de coste máximo. Sólo si estas fuesen idénticas podríamos sumarlas y obtener una nueva superficie de fricción, que utilizaríamos con la capa de direcciones de máximo coste de cualquie-

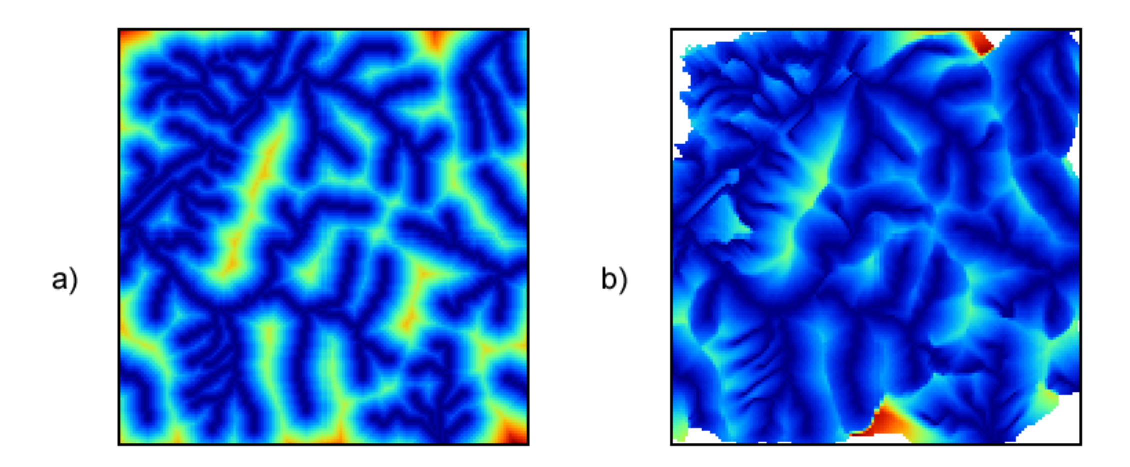


Figura 20.5: a) Distancia euclídea al cauce, b) distancia hidrológica al cauce.

ra de ellas. Este caso es altamente improbable. Incluso puede darse que en un punto el coste máximo de un factor coincida con el mínimo de otro, por ejemplo si el viento sopla pendiente arriba.

El problema estriba en la capa de direcciones, que por contener un parámetro circular tal como se vio al tratar la orientación en 16.5.1, no pueden utilizarse las operaciones aritméticas y estadísticas de la forma habitual. Este caso es similar a lo visto en 13.3.

Al igual que lo visto entonces, la forma de proceder en este caso en considerar el binomio coste–dirección (c, β) como un vector y convertirlo en sus componentes en los ejes cartesianos (x, y) según las expresiones

$$x = c \cdot \cos\beta \qquad ; \qquad y = c \cdot \sin\beta \quad (20.3.3)$$

Las componentes de cada variable de coste sobre los ejes x e y ya pueden sumarse para obtener dos capas con la componente del coste total en cada eje. Con estas dos capas puede obtenerse el par de capas resultantes con coste máximo y dirección de máxima fricción, aplicando las expresiones

$$c = \sqrt{x^2 + y^2} \qquad ; \qquad \beta = \tan\frac{y}{x} \quad (20.3.4)$$

Estas ya pueden utilizarse para obtener superficies de coste acumulado de la forma antes detallada.

20.4. Cálculo de rutas óptimas

El valor de cada celda en una capa de coste acumulado nos indica el coste mínimo para alcanzar una de las celdas de destino, pero no nos informa de la ruta que implica dicho coste. No obstante, no es difícil calcular dicha ruta.

Para entender mejor la forma de llevar esto a cabo, resulta de interés representar la capa de coste acumulado con una vista tridimensional, empleando el coste como elevación. Visualizar así esta superficie es una forma muy intuitiva de ver lo que representa y cómo utilizarla. La figura 20.6 muestra la capa de coste acumulado, así como una ruta óptima. Se ve que la superficie contiene tantos sumideros como puntos de destino. Estas son las zonas de mínimo coste (coste cero, ya que son los propios puntos de destino),

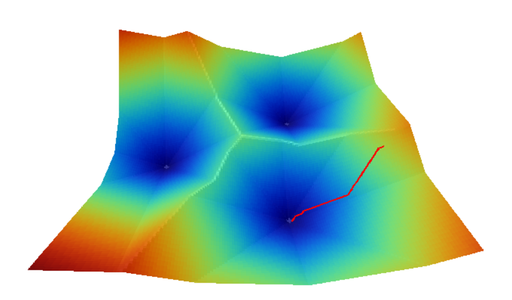


Figura 20.6: Representación tridimensional de una capa de coste acumulado y una ruta óptima sobre esta.

que aparecen con mínima elevación. La ruta va desde el punto de origen hasta el fondo de uno de dichos sumideros.

La superficie de coste acumulado es en realidad una superficie de potencial, y el desplazamiento entre el punto de origen y el de destino se asemeja mucho, como puede verse en la figura, a la ruta que seguiría un flujo desplazándose hacia aguas abajo si en lugar de coste acumulado fuera elevación el parámetro recogido en la capa. Por tanto, podemos utilizar modelos de dirección similares a los mostrados para el caso del análisis hidrológico (16.6.1). En particular, un modelo sencillo como el D8 en el que el flujo se desplaza hacia la máxima pendiente.

Por la propia forma en la que se construye la superficie de coste acumulado, no existen sumideros aparte de las propias celdas de destino, y siempre existe una celda de menor valor alrededor de cualquier otra, excepto en dichas celdas de destino, que son mínimos absolutos.

20.5. Zonas de influencia

Como ya sabemos visto, los objetos geográficos tiene influencia sobre su entorno. Un elemento lineal como un río, o uno puntual como una estación de metro, presentan una funcionalidad o un comportamiento respecto a su entorno que depende de la distancia.

A lo largo de este capítulo hemos visto que la distancia puede interpretarse como un tipo de coste. Por ello, podemos utilizar otras variables de la misma forma que la distancia para definir zonas de influencia.

Al hacerlo, podemos crear zonas de influencia de dimensión fija, tales como las creadas en forma vectorial según vimos en 19.2, o, por el contrario, de dimensión variable. Las de dimensión fija pueden tener formas irregulares alrededor del objeto central, ya que esa dimensión ya no es necesariamente en términos de distancia, sino de coste. Cumplen, no obstante, la condición de que todas las celdas en el borde de la zona tiene un mismo valor de coste (del mismo modo que, si empleamos la distancia euclídea, todos los puntos en el límite se encuentran a la misma distancia del objeto central).

En las de dimensión variable, la dimensión de la zona de influencia varía según cada celda de las que conforman el objeto, teniendo cada una una capacidad distinta de ejercer su influencia sobre el medio circundante.

Frente al cálculo de zonas de influencia que vimos para las capas vectoriales, la diferencia estriba en que en este caso no se trata de una

operación geométrica, y de que la zona no es «exacta», pues su forma y precisión depende de la resolución de celda.

20.5.1. Zonas de influencia de dimensión fija

Para comenzar, la figura 20.7 muestra una comparación entre la zona de influencia calculada sobre el trazado del cauce recogido según un modelo vectorial y la misma zona calculada sobre una base ráster. Para calcular esta última se ha creado la capa de coste acumulado tomando el cauce como conjunto de celdas de destino, pero deteniéndose el algoritmo una vez que se alcanza un umbral de distancia dado. También puede calcularse de la forma habitual, y después reclasificando todas aquellas celdas con distancia mayor que el umbral, para asignarles valor de sin datos.

La primera diferencia apreciable es que la zona de influencia en el caso ráster viene limitada a la capa de entrada en la que se contienen las entidades. En general los SIG operan de este modo, y al efectuar un proceso de álgebra de mapas la salida ráster coincide en dimensiones y tamaño de celda con la entrada. En el caso vectorial no existe restricción espacial alguna, y la zona de influencia puede «crecer» más allá de los límites de la capa de entrada.

Una segunda diferencia la encontramos en el hecho de que, además del límite de la zona de influencia, cuando esta se calcula en formato ráster existe además información en el interior de la misma. Dicha información puede servirnos para cuantificar la influencia existente dentro del área definida. Nótese en este sentido que la influencia es, por regla general, inversamente proporcional al coste, ya que cuanto más costoso sea llegar a una zona desde un punto de destino, menos influencia existirá del uno sobre el otro, así como del otro sobre el uno.

Por otra parte, este mismo proceso lo podemos realizar utilizando otras superficies de fricción, sean isotrópicas o anisotrópicas, sin estar limitados al caso de la distancia euclídea. En el caso vectorial, este cálculo no es posible desarrollarlo, ya que se trata de un proceso meramente geométrico sin el concurso de capas adicionales de fricción. Por ejemplo, la figura 20.8 muestra la misma zona de influencia anterior, con la misma distancia máxima, pero en este caso se trata de una distancia hidrológica en lugar de una euclídea.

Otro ejemplo lo encontramos en la figura 20.9, la cual representa el espacio que se supone ocupado por una determinada especie. A partir de tres puntos donde se ha detectado la presencia de dicha especie, asignando por simplicidad un coste constante a las zonas circundantes en función de su vegetación, y estimando un coste máximo a superar por un individuo de dicha especie en una jornada, se calcula la superficie de coste acumulado y se delimita la zona de influencia. Fuera de esta, es improbable encontrar individuos.

Este análisis puede realizarse de forma similar con datos vectoriales, pero en ese caso se dispone únicamente de dos clases: o el punto esta dentro de la zona de influencia o no. En el caso ráster tenemos una medida de la distancia en cada celda, que sin duda es también una medida de la probabilidad de encontrar un individuo, ya que resulta lógico pensar que en los puntos más cerca del borde la probabilidad es menor que en puntos más centrales.

Pueden calcularse también las zonas de influencia de una manera similar a la vectorial, únicamente delimitando el contorno en función de la distancia euclídea y sin aplicar los conceptos de creación de capas de coste acumulado. En este caso basta tomar todas aquellas celdas que consti-

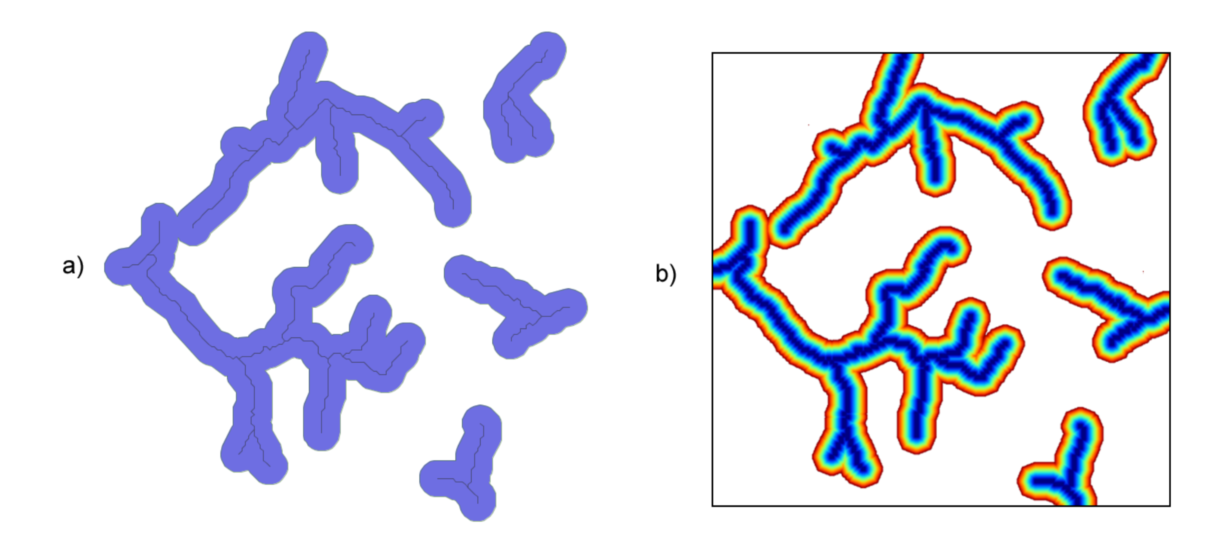


Figura 20.7: Comparación entre a) zona de influencia en formato vectorial y b) zona de influencia en formato ráster.

tuyen los objetos centrales (las celdas de destino en el caso del análisis de coste) y marcar con un valor establecido las celdas circundantes a una distancia menor que la distancia de influencia escogida. Se trataría de un análisis focal con una ventana de análisis circular de radio igual a la distancia de influencia, en la que los valores dentro de esta reciben todos el mismo valor. Obviamente, los resultados que pueden obtenerse de este modo son más limitados que aplicando toda la potencia del análisis de costes.

20.5.2. Zonas de influencia de dimensión variable

Al realizar el cálculo de una zona de influencia de dimensión fija, establecemos un umbral de coste acumulado, a partir del cual consideramos que no existe tal influencia. Todos los elementos en el limite del área delimitada tienen el mismo valor de coste acumulado. En ocasiones, no obstante, lo interesante para definir la influencia de un elemento geográfico puede no ser el coste acumulado desde el mismo a las celdas del entorno, sino el propio coste unitario de dichas celdas, o bien otro valor asociado a las mismas, no necesariamente un coste.

Con estos planteamientos podemos definir zonas de influencia de dimensión variable, en las cuales las celdas fronterizas no cumplen ningún requisito relativo al coste acumulado que se da en las mismas.

Para ver un primer ejemplo considérese el siguiente supuesto: se dispone de una carretera y una capa de pendientes. Por la carretera los vehículos circulan sin dificultad, y fuera de ella, los vehículos todo-terreno pueden hacerlo siempre que la pendiente no sea superior al 5 %. Tratemos de calcular la zona de influencia de la carretera, es decir, la zona que es accesible con un vehículo todo-terreno.

Podemos abordar el problema como un problema de costes habitual. Tomando la superficie de fricción, reclasificamos todos los valores por encima de nuestro umbral del 5 % y les asignamos

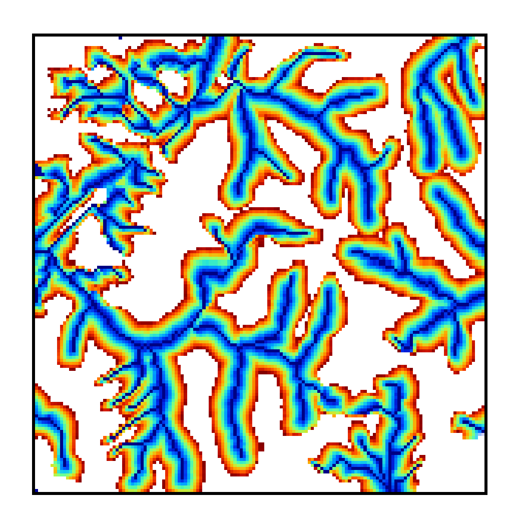


Figura 20.8: Zona de influencia de un cauce basada en distancia hidrológica.

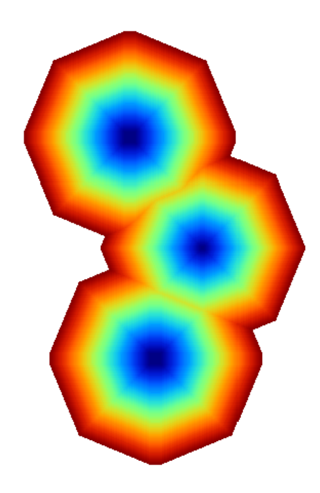


Figura 20.9: Zona de influencia de una especie dada su área de residencia y un coste máximo de desplazamiento.

valor de sin datos para indicar que no son transitables. Despues, calculamos la superficie de coste acumulado, tomando las celdas de carretera como celdas de destino (Figura 20.10). No todas las celdas con pendiente inferior al 5 % forman parte del área de influencia, ya que, aunque el vehículo puede transitarlas, algunas no puede alcanzarlas, y quedan como «islas».

El modelo de coste, pese a incluir la pendiente, es en esta ocasión isotrópico, ya que el vehículo no puede desplazarse por zonas con pendiente superior al umbral, con independencia de la dirección en la que lo haga.

Asimismo, no es necesario aplicar ningún umbral a esta capa de coste acumulado, ya que no es ese parámetro el que define la zona de influencia. Si la carretera esta rodeada a ambos lados por zonas completamente llanas, la zona de influencia se extenderá indefinidamente, ya que el coste acumulado no es relevante en este caso. Lo utilizamos simplemente para incorporar la conectividad de las distintas celdas transitables con la vía central. Es por ello que el mapa de la figura 20.10 solo marca la zona de influencia sin incorporar los valores interiores de coste acumulado (en este caso pendiente acumulada), ya que no son de interés.

Podemos añadir más complejidad al modelo utilizando un umbral variable. Por ejemplo, sea un cauce del cual conocemos, en cada una de sus celdas el valor de su calado en un evento extremo. El cauce viene definido como una estructura lineal de una única celda de ancho, pero con estos datos vamos a tratar de definir el área realmente ocupada por el agua en ese evento. Esto es, el área de inundación.

Modelizar hidráulicamente un cauce en un supuesto como el presentado es mucho más complejo que lo que vamos a ver, y se requieren más datos, pero podemos plantear una primera aproximación al estudio de ese área de inundación, pues no deja de ser una zona de influencia.

Si en el caso del vehículo teníamos un umbral fijo, ahora este umbral es variable y depende del calado. Como muestra la figura 20.11, para una elevación z en la celda de cauce y un calado h, el agua podrá inundar aquellas celdas contiguas con elevación menor que $z + h$. Si tomamos la

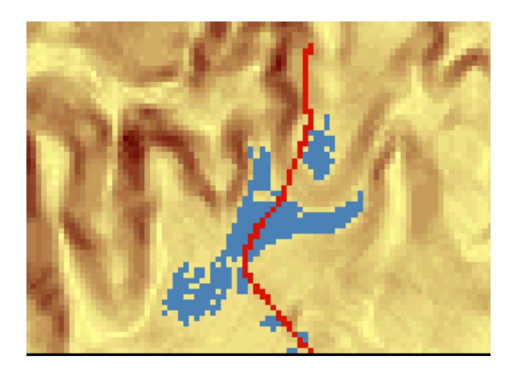


Figura 20.10: Zona de influencia (en azul) de una vía (en rojo) , suponiendo que es posible el desplazamiento desde esta siempre que la pendiente no supere el 5 %. En el fondo, mapa de pendientes. Tonalidades más oscuras indican mayor pendiente.

capa de calado y le sumamos el MDE, tendremos una capa de cauces en las que sus celdas representan los puntos de destino[1], y además cada una de ellas contiene el valor de umbral. Es decir, que al operar según se explicó anteriormente para crear la capa de coste acumulado, el umbral dependerá de la celda concreta desde la que nos venimos desplazando. Las celdas por debajo del umbral son viables, mientras que las superiores, no. Este modelo es similar al que planteábamos al analizar la distancia hidrológica, solo que en este caso el umbral que aplicamos no es sobre el coste acumulado, sino que lo usamos para calcular en cada celda los costes unitarios. Para aquellas celdas que superan dicho umbral, el coste es infinito. Para las restantes, nulo.

20.6. Influencia acumulada

En los anteriores ejemplos, delimitamos la influencia de un elemento geográfico a un entorno de este, y en algunos casos cuantificamos esta. Si el objeto geográfico se compone de varias celdas, por la propia naturaleza del algoritmo, solo será la más cercana la que tenga influencia en el valor final de coste acumulado de otra celda dada. Esto, que resulta correcto para el cálculo de rutas óptimas, no lo es tanto cuando queremos considerar simultáneamente todas las celdas de destino (o de influencia, según sea el enfoque).

Supongamos que existen núcleos de población cercanos al cauce, cada uno de los cuales tiene una población dada. Si desde estos núcleos la población va hacia el río a practicar la pesca, la influencia de esta actividad sobre cada celda del cauce dependerá de todas las poblaciones cercanas. Dependerá, asimismo, de la distancia o coste de desplazamiento desde las mismas y de la población de estas.

Calcular una capa de coste acumulado utilizando todas las ciudades no sirve en este caso para evaluar su influencia sobre las distintos tramos del cauce. Si tomamos un coste máximo, podemos delimitar una zona de influencia y ver qué zonas del cauce no se ven afectadas. Para las restantes, sin embargo, la influencia es distinta, y

[1] Si asumimos que las celdas sin calado (aquellas que estén fuera de cauce) tienen un valor de sin datos, al sumar las dos capas, y como se dijo en 15.2.2, todas estas celdas tendrán valor de sin datos en la capa resultante, independientemente del valor de elevación que tengan en el MDE.

Por ello, esta nueva capa también nos servirá como capa puntos de destino según la codificación habitual

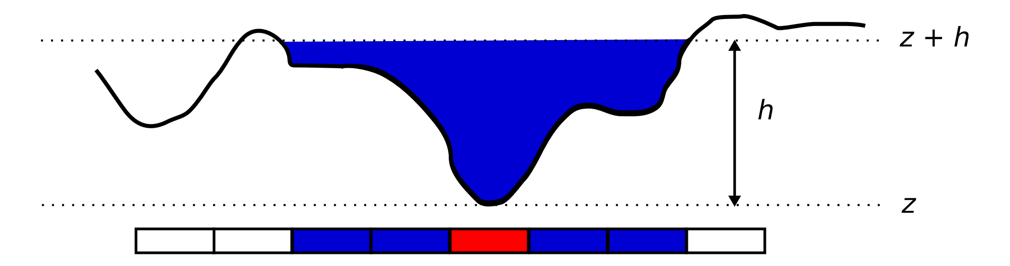


Figura 20.11: El calado h define la inundabilidad de las zonas circundantes al cauce, en función de la elevación de estas. En rojo, celda central del cauce.

la información de que disponemos en esa capa de coste acumulado no refleja la verdadera influencia del conjunto de ciudades. No obstante, los elementos del análisis de costes sí que nos sirven para, aplicándolos de un modo distinto, obtener el resultado buscado.

En 14.3 suponíamos que la zona de influencia para el cálculo de densidades era circular, llevando esto implícito que el coste utilizado era la distancia. Ahora conocemos el modo de emplear otras variables de coste y, siendo similar el cálculo de densidades (o probabilidades) al de influencias, podemos proceder de forma similar a lo explicado entonces, operando con las capas de costes asociadas a cada punto de influencia. He aquí la manera de hacerlo:

- Calculamos la superficie de coste acumulado empleando una única ciudad como destino.
- Con la capa anterior, creamos una capa de influencia de dicha ciudad sobre su entorno. La influencia en este caso es inversamente proporcional al coste, así que podemos sencillamente invertir sus valores o aplicar otra función más compleja que relacione ambas variables.
- Repetimos el anterior proceso para cada una de las ciudades. Obtendremos tantas capas de influencia como ciudades existan.
- Sumamos las capas anteriores para obtener la influencia del conjunto de ciudades en cada celda.
- Puesto que estamos estudiando la influencia de la práctica de la pesca, aplicamos una máscara para que la capa anterior quede restringida a las zonas donde esta pueda desarrollarse, es decir, los cauces.

20.7. Análisis de redes

Todo el análisis de costes lo hemos realizado hasta este punto sobre una base ráster. Esta es la manera más ventajosa y adecuada de llevarlo a cabo, pues trabajamos con superficies de tipo continuo (las superficies de fricción), que se analizan mejor bajo este modelo de representación.

Un caso particular del análisis de costes es el análisis del movimiento a través de una red, siendo el caso más habitual en un SIG el de una red viaria. El calculo de la ruta de menor coste entre dos nodos de dicha red es un problema típico que tiene numerosas aplicaciones incluso fuera del ámbito geográfico, y constituye una aproximación diferente a la que ya conocemos para el caso ráster.

En el supuesto de un desplazamiento por una red viaria, la división del terreno en unidades regulares que implica el modelo ráster no es adecuada, y resulta mucho más lógico un modelo vectorial que contenga las vías y los emplazamientos entre los que estas se sitúan, recogiendo igualmente la topología de la red. Esta estructu-

ra es óptima no solo para el almacenamiento y manejo, sino también para el análisis de costes, como a continuación veremos.

El análisis de costes ráster tal como lo hemos visto permite modelizar los costes a través de vías y el movimiento restringido a estas. Basta excluir las celdas fuera de las vías (asignándoles coste infinito o valores de sin datos) y operar de la forma habitual. No obstante, el hecho de que exista dicha restricción (al analizar un coste de desplazamiento por vías, no podemos abandonar estas), favorece la utilización de un modelo vectorial para estos casos, haciendo que resulte más lógico. Otros elementos de la vía, como el hecho de que algunas de ellas puedan recorrerse únicamente en una dirección, se recogen mejor mediante un modelo de representación vectorial, como ya vimos en el capítulo 5 al introducir el concepto de topología (de ese capítulo deberías recordar también los conceptos de arco y nodo, que aquí resultan básicos).

Este es el tipo de modelo que implementan, por ejemplo, los navegadores GPS, ya que el movimiento que estudian y sobre el cual nos informan se realiza siempre por una red de carreteras y calles. Sin embargo, si lo que pretendemos es calcular un coste de desplazamiento no confinado a una vía (por ejemplo, para calcular la ruta de menor coste caminando campo a través), es el modelo ráster el que debemos utilizar.

Aunque formalmente son distintos tanto los modelos de representación como los casos particulares de análisis, la implementación de los algoritmos de cálculo de rutas de mínimo coste es en cierta medida similar. Así, los algoritmos sobre una base vectorial comparten elementos con los que ya conocemos para el análisis ráster. Estos algoritmos en realidad derivan todos ellos de la denominada *teoría de grafos*, una rama de la matemática que se encarga de trabajar con un tipo de estructuras denominadas *grafos*, de los cuales una red es una caso particular.

El cálculo de rutas óptimas es una de las áreas más trabajadas dentro de la teoría de grafos, y por ello existen muchos algoritmos distintos. El más popular de todos ellos es el conocido como algoritmo de Dijkstra, que es en cierto modo en el que se basan las metodologías que hemos visto para el caso ráster. En realidad, las celdas de una capa ráster pueden entenderse también como una estructura de nodos conectados (una red), con la particularidad de que estas conexiones tienen un carácter sistemático: cada celda (nodo) está conectado siempre a las ocho que se sitúan entorno a ella, y únicamente a estas.

En el algoritmo de Dijkstra, se inicia el cálculo en el nodo de destino y se *marcan* los nodos a los que este se encuentra conectado, asociándoles el valor de coste entre dicho nodo inicial y cada uno de estos otros nodos. Los restantes nodos, hasta que no sean marcados con un valor concreto de coste, tienen un coste infinito, pues aún no se ha establecido una manera de conectarlos con el nodo de destino.

Al acabar este proceso, se procede a repetirlo, en esta ocasión con todos los puntos que han sido marcados en la iteración anterior. En esta nueva iteración, cada uno de los nodos conectados recibe como nuevo valor de coste el del nodo base (que será uno de los conectados al nodo original) más el coste de desplazarse entre este nodo y el nodo conectado a él. En caso de ya tener un valor de coste asociado, solamente se le asociara este nuevo si es inferior al valor existente (es decir, solo si la ruta que estamos evaluando en este paso es de menor coste que la que anteriormente se evaluó al asignar dicho valor de coste previo). Este proceso de recalculo de coste asociado a un nodo se conoce como *relajación*.

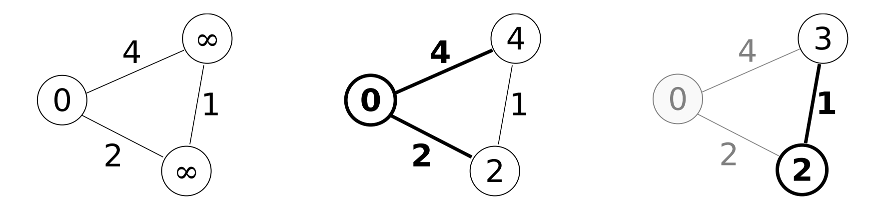


Figura 20.12: Ejemplo de aplicación del algoritmo de Dijkstra para cálculo de rutas de mínimo coste. De izquierda a derecha, etapas sucesivas de asignación de coste por nodos (Adaptado de Wikipedia).

La figura 20.12 muestra un ejemplo de un grafo muy sencillo y el proceso de ejecución del algoritmo de Dijkstra sobre este.

La descripción original del algoritmo puede consultarse en [282]. Para el lector interesado en profundizar sobre esta materia, un buen compendio de algoritmos de cálculo de rutas óptimas puede encontrarse en [283].

Existen numerosos cálculos relacionados con las redes y la teoría de grafos cuya importancia dentro del ámbito SIG es notable. Uno de los más conocidos es el problema del árbol mínimo de recubrimiento, más habitualmente denotado por su nombre en inglés: *minimum spanning tree* (MST). Dado un conjunto de puntos, este árbol representa el mínimo conjunto de líneas que permite conectar todos esos puntos. Es decir, el problema del MST implica crear una red que aporte conectividad a un conjunto dado de puntos, teniendo esa red la mínima longitud posible.

Este problema clásico en la teoría de grafos (y por tanto también muy estudiado al igual que los relativos al cálculo de rutas mínimas que acabamos de ver) tiene numerosas aplicaciones cuando se traslada al campo del análisis geográfico. Por ejemplo, permite calcular la red de canalizaciones necesaria para abastecer a una serie de puntos, minimizando el gasto en tuberías.

La figura 20.13 presenta un ejemplo de uno de estos árboles.

El problema puede resolverse considerando distancia euclídea, o bien teniendo en cuenta que los puntos se encuentran conectados por una red, con un coste dado entre cada par de ellos. En este segundo caso, la obtención del MST implica la reducción de la red original que los conecta, eliminando tramos hasta lograr el conjunto mínimo de ellos que mantiene la conectividad.

De entre los múltiples algoritmos existentes para resolver este problema, los de Prim[284] y Kruskal[285] son los más habituales. Los propuestos originalmente por el matemático checo Otakar Boruvka constituyen la base para gran parte de las formulaciones más elaboradas. Pueden encontrarse en [286]

Puesto que se obtiene como resultado un árbol y este es un grafo de tipo acíclico, la ausencia de ciclos (circuitos cerrados) garantiza que no existan tramos «redundantes» en la red. Por esta razón, la estructura de árbol es la adecuada para minimizar la longitud de la red. Desde la perspectiva de su aplicación real, no obstante, un árbol no es la forma más ventajosa de conectar una serie de puntos, ya que la conectividad es reducida y puede perderse si se pierde una de las conexiones (en otras palabras, si se rompe una tubería, habría puntos que no estarían abastecidos, y no resultaría posible abastecerles por otra vía).

Una solución más adecuada es optar por una estructura que trate de reducir la longitud total de la red, pero garantizando una conectividad más robusta. El MST guarda mucha relación con una estructura que ya conocemos, la triangulación

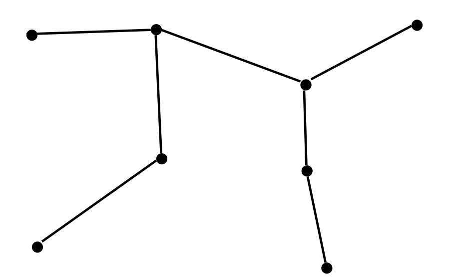


Figura 20.13: Árbol mínimo de recubrimiento para un conjunto de puntos.

Figura 20.14: Solución al problema del viajante para un conjunto de puntos.

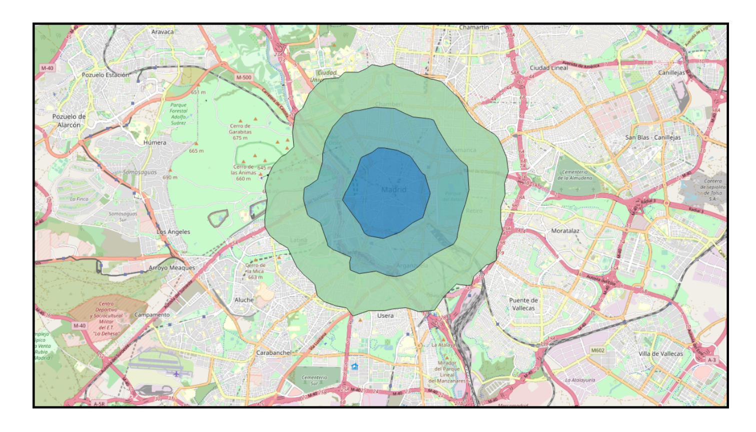


Figura 20.15: Zona de influencia de un punto considerando distancias sobre una red en lugar de distancia euclídea.

de Delaunay, ya que se forma como un subconjunto de las líneas que conforman esta (puede consultarse por ejemplo[287] para ver detalles acerca del algoritmo de cálculo del MST a partir de la triangulación). Si de ese conjunto total de la triangulación se toman las líneas del MST y algunas adicionales, pueden obtenerse estructuras que solucionan de manera óptima el problema de conectar un conjunto de puntos con un diseño de red robusto. Un ejemplo de ésto son las denominadas redes de Gabriel [288].

También en cierta forma relacionado con los problemas anteriores, un enunciado clásico con gran aplicación en el ámbito SIG es el conocido como *problema del viajante* o TSP[2]. Dado un conjunto de puntos, se trata de calcular la forma de visitar todos ellos en un orden dado y regresando al punto inicial, de tal modo que el recorrido total tenga la mínima longitud posible.

La forma más directa y simple de resolver este problema es probar todas las posibles rutas y elegir la más corta. Sin embargo, debido al crecimiento exponencial del número de posibilidades, este planteamiento es inviable, y se ha de optar por métodos de resolución aproximada. Dado el grado de estudio este problema, estas soluciones son muy variadas, y existe bibliografía muy abundante al respecto.

En relación con los SIG y, particularmente, con el tema de este apartado, el problema del via-

[2] *Travelling Salesman Problem*

jante cobra interés cuando se considera, al igual que ya veíamos para el MST, que las distancias entre los distintos puntos no han de ser necesariamente euclídeas, sino que pueden ser distancias a través de una red. Así, si una persona ha de visitar un número dado de emplazamientos dentro de una ciudad desplazándose en un automóvil, el orden óptimo en que debe hacerlo es el que se obtendría como solución al problema del viajante para dichas localizaciones, y tomando las distancias entre estas como distancias por las calles de la propia ciudad.

Independientemente de la métrica empleada para calcular la distancia entre puntos, la solución del problema se puede realizar empleando la misma metodología.

En la figura 20.14 puede verse el circuito óptimo para el conjunto de puntos empleado en el ejemplo anterior de la figura 20.13.

Para concluir este apartado, comentar que el concepto de *buffer* vectorial puede extenderse si disponemos de una red sobre la que calcular distancias, haciéndolos más similares a los que hemos visto para el caso ráster, donde la anchura de estos era variable y no se presentaba la característica simetría de los que vimos en el capítulo 19. Sobre dicha red, podemos calcular puntos a una distancia dada, y con ellos crear el polígono que delimita la zona de influencia.

La figura 20.15 muestra una red viaria con costes asociados, y una posible zona de influencia basada en dichos costes en lugar de en distancia euclídea.

20.8. Resumen

Las superficies de fricción contienen valores de coste que expresan la resistencia que presenta una celda a ser recorrida. Estos costes pueden reflejar muchos distintos factores, siendo la distancia uno de ellos.

Con una superficie de fricción y un conjunto de puntos de destino, se calculan capas de coste acumulado. Empleando estas es posible calcular rutas óptimas que definen la ruta de menor coste desde un punto dado hasta el punto de destino más cercano (en términos de coste).

Con estas ideas, pueden definirse zonas de influencia tanto de dimensión fija como de dimensión variable, y puede estudiarse asimismo la influencia conjunta de una serie de elementos geográficos sobre el entorno inmediato de estos.

Aunque el análisis de costes y superficies de fricción es un análisis con elementos ráster, las redes vectoriales con topología permiten un análisis distinto para calcular rutas óptimas entre puntos de dichas redes.

Capítulo 21

Más estadística espacial

21.1. Introducción

Prácticamente todas las formulaciones estadísticas pueden aplicarse a datos geográficos de igual forma que se aplican a datos de otro tipo. No obstante, la particular naturaleza de la información geográfica hace que algunas de estas formulaciones puedan adaptarse de forma específica para obtener resultados más correctos. En particular, la inferencia estadística presenta, como ya vimos, una serie de peculiaridades que hacen conveniente su adaptación para considerar la especial naturaleza debida a la presencia de una componente espacial.

En este capítulo, intentaremos profundizar algo más sobre el empleo de estas herramientas estadísticas dentro del ámbito de un SIG, viendo, entre otros, los métodos de regresión que consideran explícitamente la componente espacial de los datos, así como las metodologías que constituyen la base para la creación de modelos predictivos, de gran aplicación dentro de los SIG.

Asimismo, existe una operación estadística con un particular peso dentro del ámbito del SIG: la clasificación, la cual constituye un elemento básico especialmente en el tratamiento de imágenes y la creación de datos espaciales de carácter temático a partir de estas. Dedicaremos gran parte del capítulo a detallar los pormenores del proceso de clasificación y algunas, las más populares, de sus muy diversas variantes.

Ya conocemos algunas formulaciones que nos permiten, a partir de una capa ráster (sea una imagen o no), obtener capas vectoriales o nuevas capas ráster de variable discreta, conteniendo una información adecuada para ser expresada en formato vectorial, de forma similar a las operaciones de clasificación. El capítulo dedicado al procesado de imágenes contenía algunas de estas formulaciones, basadas en aquel caso en operaciones con apenas componente estadística. Los procesos de clasificación de este capítulo abordan, como hemos dicho, una tarea similar, pero desde un enfoque principalmente estadístico.

El *Análisis de Componentes Principales*, otra técnica estadística habitual, también es de gran utilidad en el trabajo con imágenes, especialmente cuando estas contienen un gran número de bandas. La redundancia que aparece en estas imágenes al existir dependencia entre sus bandas puede eliminarse aplicando esta técnica, al igual que cuando se trabaja con cualquier otro conjunto de múltiples capas, cada una de ellas referida a una variable distinta.

Por último, cerraremos el capítulo presentando algunas metodologías que nos ayudaran a extender las capacidades de las operaciones locales del álgebra de mapas, es decir, aquellas que empleábamos para combinar una serie de capas ráster. Las operaciones algebraicas que utilizamos para efectuar dicha combinación de capas pueden mejorarse si se aplica un análisis estadís-

tico de las distintas variables combinadas, dando esto lugar a metodologías más avanzadas.

21.2. Clasificación

La clasificación es una operación muy habitual dentro del análisis SIG, pues permite establecer una categorización de un área de estudio a partir de una serie de variables, con la utilidad que ello conlleva para otros análisis. Esencialmente, el proceso de clasificación convierte una serie de capas en otra única con información categórica. Las capas de origen suelen contener en general variables de tipo continuo, aunque no necesariamente. Se trata de, en función de la información disponible sobre un área geográfica, clasificar de acuerdo con algún criterio las distintas zonas de la misma.

Esta clasificación se desarrolla mediante procedimientos diversos que evalúan a cuál de las clases posibles es más similar un punto dado. En dicho punto, se toman los valores de todas las variables registradas y ese conjunto de valores es el utilizado para calcular el grado de similitud entre la clase presente en el punto (aún no determinada) y las distintas clases posibles.

Dos circunstancias son de reseñar en relación con la naturaleza de las capas de partida:

- Por tratarse de variables continuas en la mayoría de los casos, se suele trabajar principalmente con capas ráster como capas de origen.
- Por utilizarse en general más de una variable, se requieren varias capas para almacenar estas.

Como ya vimos, la estructura de las imágenes procedentes de teledetección, en formato ráster y con múltiples bandas, las hace ideales para ser utilizadas como datos de partida para obtener capas clasificadas. Por esta razón, es frecuente ver los algoritmos de clasificación tratados en textos dedicados al análisis de imágenes, o bien en las partes sobre imágenes de los textos genéricos sobre SIG. No obstante, no deben entenderse como limitados a este tipo de datos, ya que resultan igualmente de utilidad aplicados sobre cualquier conjunto de capas ráster, sin que estas sean bandas de una imagen o sin que las variables empleadas hayan sido recogidas por un mismo instrumento o sensor.

Un número dado de valores recogidos en otras tantas bandas de una imagen sirven para evaluar, por ejemplo, el tipo de material del que se compone un suelo, puesto que las distintas clases establecidas en función de esos materiales guardan relación con las intensidades a distintas longitudes de onda que dicho suelo refleja. Pero del mismo modo, y como ya se comentó en 16.8, las formas del relieve también pueden dividirse en clases atendiendo a la pendiente, la curvatura o el índice de humedad topográfica, entre otros parámetros. Y de manera similar, establecer una caracterización de las distintas zonas puede llevarse a cabo a partir de valores de pH del suelo, pendiente o humedad, por nombrar algunos.

Incluso pueden combinarse bandas procedentes de imágenes y otro tipo de variables. Por ejemplo, al clasificar las distintas zonas de vegetación a partir de una imagen, pueden añadirse valores de pendiente o elevación, ya que estos parámetros del terreno tienen una obvia relación con la vegetación existente.

Como veíamos en el capítulo sobre imágenes, el conjunto de valores de las distintas bandas de una imagen en un punto constituía lo que denominábamos *firma espectral*. Al trabajar con otro tipo de variables, el conjunto de valores en un punto que empleamos para clasificarlo ya no recibe ese nombre, que es propio del análisis de imágenes, pero el concepto sigue siendo, sin embargo, el mismo.

En el caso de imágenes, la clasificación es, además de una técnica estadística, un proceso

con una base física, pues pueden estudiarse directamente los valores de radiación reflejada a distintas longitudes de onda y adjudicarles un significado en función de lo que dichas longitudes de onda condiciones. En otros ejemplos, sin embargo, se trata de un proceso meramente estadístico. Esta componente estadística y su formulación será lo que veamos en este apartado.

La clasificación, pues, puede definirse como el proceso que, dados un conjunto de elementos (en este caso, localizaciones espaciales) las agrupa en una serie de clases de forma que estas sean homogéneas en cuanto a las características de los elementos que contienen.

Este proceso conlleva dos etapas:

- Definición de las clases
- Asignación de cada elemento a una de dichas clases

En función de cómo se lleve a cabo la definición de las clases, los métodos de clasificación pueden dividirse en dos grupos principales:

- Clasificación supervisada. Además de emplear las capas como entrada, debe añadirse información adicional que ayude en la definición de las distintas clases, definiendo el número de estas y sus características.
- Clasificación no supervisada. La única entrada son las capas y el número de clases a definir. Las características de dichas clases se establecen en función del conjunto de valores con los que se trabaja.

21.2.1. Clasificación supervisada

La clasificación supervisada es una forma de clasificación que requiere por parte del operador la definición explicita de las clases a definir. En la terminología empleada para las imágenes, podemos decir que el operador debe introducir la firma espectral característica de las clases, expresada esta como los valores más habituales que aparecen para dicha clase. El proceso de clasificación asigna a un punto aquella clase cuyo conjunto de valores «típicos» de las variables estudiadas (los que vienen definidos por esa firma espectral característica) son más similares a los presentes en dicho punto.

La figura 21.1 muestra un esquema de este proceso.

La forma en que pueden proporcionarse definiciones de clase a un método de clasificación supervisada son dos:

- Mediante zonas de entrenamiento
- Mediante valores directos

Al utilizar zonas de entrenamiento, el operador debe delimitar algunas zonas en las que se conozca *a priori* el tipo de clase existente. Esto puede hacerse con una capa de polígonos adicional que tenga asociado en un campo de su tabla de atributos el tipo de clase presente en dicho polígono. El algoritmo de clasificación, en su primera fase, estudiará las celdas de las capas de entrada que caen dentro de los polígonos de cada clase, y con sus valores tratará de definir los rasgos generales de esas clases que permitirán identificar clases similares en otros puntos.

A la hora de definir estas zonas de entrenamiento, debe procurarse que cubran toda la casuística de las clases que definen. Si, por ejemplo, queremos clasificar un área de estudio en tres simples clases como «bosque», «cultivo» y «carretera», y las zonas de bosque son heterogéneas (distintas especies, distinta densidad, etc.), será conveniente definir zonas de entrenamiento en distintos tipos de bosque, para que puedan extraerse las características comunes a todas las subtipologías que vamos a englobar en una única clase.

Por supuesto, si queremos definir una clase concreta, debemos establecer una o varias zonas de entrenamiento de esta. Es decir, el algoritmo solo clasifica en los grupos que el operador ha-

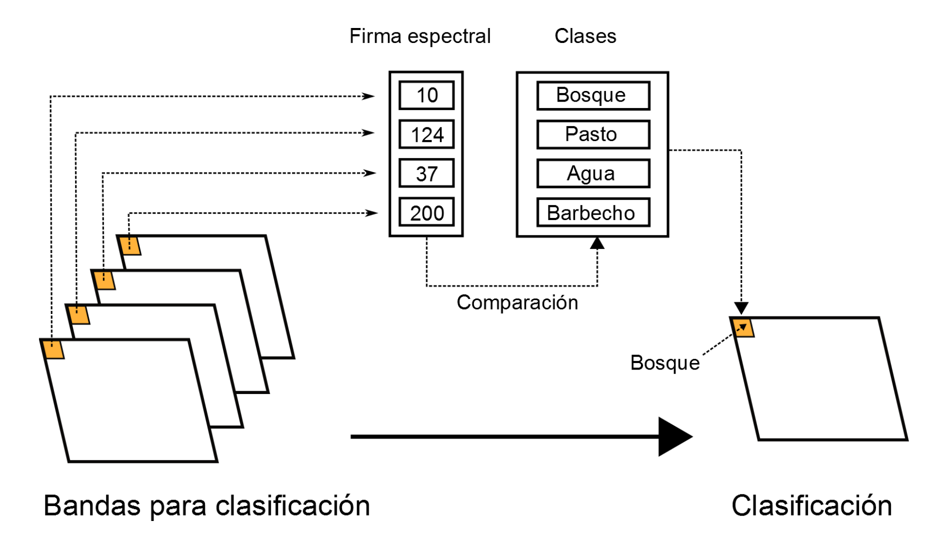


Figura 21.1: Esquema del proceso de clasificación supervisada

ya definido. Si en el área de estudio mencionada anteriormente existieran zonas en barbecho (que no pertenecen a ninguna de las tres clases señaladas), estas no van a asignarse a una clase nueva. En función de la metodología que posteriormente se use para dicha asignación, o bien quedarán sin clasificar (si son muy distintas sus características de las de todas las clases establecidas), o bien quedarán englobadas dentro de la clase con la cual presenten una mayor similitud (que probablemente, y pese a ser la más similar de las tres definidas, no sea muy parecida a la realidad).

Para obtener mejores resultados, las zonas de entrenamiento deben ser lo más representativas posibles, y deben establecerse siempre que se tenga la seguridad de que efectivamente pertenecen a la clase a definir. Un mayor número de zonas de entrenamiento implica mejor clasificación en líneas generales, pero solo si estas zonas constituyen una definición coherente de la clase, y no si se emplean zonas cuya tipología no se conozca con plena certeza.

Cuanto más grandes sean las zonas de entrenamiento, más celdas contendrán en las capas de variables analizadas, y más precisa sera la definición de las características de cada clase. [289] recomienda que se tomen para cada clase al menos un número de celdas diez veces superior al de variables consideradas, siendo más adecuado que este sea cien veces mayor.

Además de emplear zonas de entrenamiento, la otra forma de comunicarle al algoritmo de clasificación qué clases queremos definir y qué características tienen estas es mediante valores directos. En el análisis de las zonas de entrenamiento, las zonas se caracterizan según unos valores estadísticos tales como la media y la desviación típica de las distintas variables empleadas. Si ya hemos realizado ese proceso anteriormente, dichos valores ya han sido calculados, y podemos utilizarlos para alimentar un nuevo proceso de clasificación en otro área de estudio. Lógicamente, esto es posible solo si las clases a definir tienen características similares en este nuevo área y en la utilizada originalmente para la caracterización de clases.

La utilización de valores provenientes de bases de datos de firmas espectrales, las cuales se

vieron en el capítulo 17, constituye otro ejemplo del uso de valores directos en lugar de zonas de entrenamiento.

Una vez que las clases se han definido, el proceso de clasificación asocia cada punto de la zona de estudio a una de ellas en función de sus valores. Los métodos existentes para llevar a cabo esta fase del proceso son muy diversos, y la literatura estadística recoge decenas de ellos. Dentro del ámbito de los SIG, los más habituales son los siguientes:

- Clasificación por paralelepípedos.
- Clasificación por mínima distancia.
- Clasificación por máxima verosimilitud.

Se trata de métodos sencillos en su mayoría, existiendo formulaciones más complejas y avanzadas que, no obstante, resulta menos frecuente encontrar implementadas en un SIG. Algunos de estos métodos que merecen ser mencionados son los basados en redes neuronales, árboles de decisión o sistemas expertos. La ventaja frente a los anteriores es que no asumen una distribución estadística particular de los datos, y pueden aplicarse aun en el caso de que dicha distribución no se dé.

Para más información, pueden consultarse, entre otras referencias, [290], [291], [292] y [293]. Métodos basados en lógica difusa se recogen, por ejemplo, en [294]. Todas estas metodologías se presentan generalmente aplicadas a la clasificación de imágenes, aunque una vez más su uso no se ha de restringir al caso particular de estas.

Antes de aplicar cualquiera de estos métodos o los que seguidamente veremos con detalle, puede resultar recomendable homogeneizar los rangos de las distintas variables. En el caso de emplear únicamente imágenes, los Niveles Digitales de estas se encuentran siempre en el mismo rango (0–255), pero este puede ser más heterogéneo si se usan capas con otro tipo de variables. Un proceso de normalización (lo vimos en 15.2.1) es una opción habitual en este caso, empleándose como preparación previa al análisis y la clasificación.

Paralelepípedos

El método de clasificación por paralelepípedos establece regiones, una por cada clase, con dicha forma de paralelepípedos dentro del espacio de atributos en el que se trabaja. La pertenencia de una localización a una de las clases se establece viendo si la posición que sus valores definen en el espacio de atributos está dentro de la región correspondiente a la clase.

Los distintos lados de los paralelepípedos vienen definidos por la media de cada variable en la clase y su desviación típica. La longitud de los lados es igual a dos veces la desviación típica, y se encuentran centrados en la media. Un esquema de esto puede verse en la figura 21.2. Los puntos representados son elementos que se quieren clasificar, que en el caso de una capa ráster serán mucho más numerosos, uno por cada celda. Por simplicidad de representación y visualización, se supone en dicha figura, así como en las siguientes correspondientes a otros métodos, que se utilizan dos variables para efectuar la clasificación (por tanto, es un gráfico bidimensional), y se definen tres únicas clases. El mismo tipo de análisis puede, no obstante, aplicarse a cualquier numero n de variables (resultaría un espacio de atributos n–dimensional) y, por supuesto, cualquier número de clases.

Una característica de este método es que pueden existir elementos que no puedan ser clasificados, al no caer dentro de ningún paralelepípedo. Del mismo modo, estos pueden solaparse y ciertos elementos pueden pertenecer simultáneamente a varias clases. En caso de darse esta circunstancia, puede resolverse aplicando alguno de los otros métodos, tal como el de máxima verosimilitud.

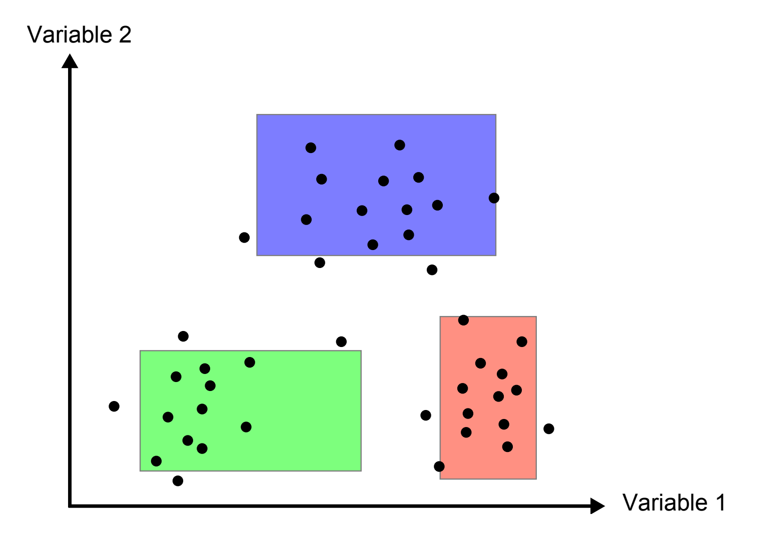


Figura 21.2: Método de clasificación por paralelepípedos

En general, la precisión de este método es baja, y el número de elementos sin clasificar o clasificados en varias categorías es alto. Su mayor ventaja reside no en su exactitud, sino en la velocidad de proceso, al no requerir operaciones complejas.

Mínima distancia

El método de mínima distancia se basa también en conceptos geométricos dentro del espacio de atributos. En este caso se emplea únicamente la media de cada clase, prescindiéndose de la desviación típica.

Para cada uno de los elementos a clasificar se calcula la distancia euclídea en el espacio de atributos entre la media de cada clase y dicho elemento. Esta distancia viene expresada por

$$d_k = \sqrt{(x_i - \overline{x}_{ik})^2} \qquad (21.2.1)$$

siendo d_k la distancia del elemento al centro de la clase k-ésima, x_i el valor asociado al elemento para la variable i-ésima, y $\overline{x}_{ik}$ la media de los valores de la clase k-ésima para la variable i-ésima.

Aquella clase hasta la que exista una menor distancia será a la que se asigne el elemento en cuestión.

En lugar de emplear distancia euclídea puede utilizarse la distancia de Manhattan, ya que disminuye el número de operaciones (lo cual implica más velocidad de proceso).

Puede verse que esta metodología guarda similitud conceptual con la interpolación por vecindad, en la que asignábamos el valor del punto más cercano. En la figura 21.3 puede verse un gráfico explicativo de este método de clasificación.

A diferencia del método anterior, todos los elementos pueden clasificarse, ya que siempre existe uno más cercano. Esto constituye una de las potenciales desventajas del método, ya que puede hacer asignaciones incorrectas en el caso de que un elemento sea muy distinto a todas las clases de partida. Se asignará a la clase más similar, lo cual no significa necesariamente que sea lo suficientemente similar a ella como para considerarlo parte de la misma.

Por ejemplo, y repitiendo un ejemplo ya citado, si tenemos las clases «bosque», «cultivo» y

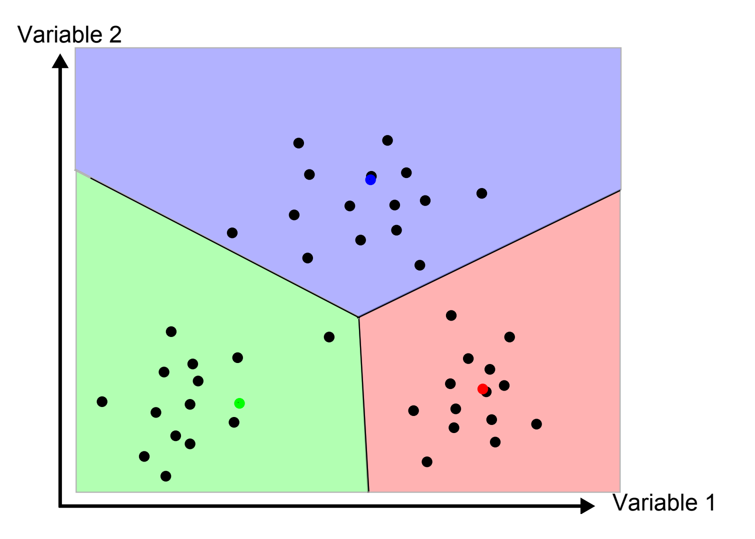


Figura 21.3: Método de clasificación por mínima distancia

«carretera», y en nuestro área de estudio existe una zona en barbecho, esta será clasificada dentro de alguno de los grupos anteriores, que puede ser uno u otro en función de los datos que empleemos para la clasificación. Con independencia de cuál sea esa clase escogida, resulta claro que sera una asignación errónea, y que o bien debería haberse incorporado esta clase dentro de las zonas de entrenamiento, o bien toda esa zona de barbecho debería quedar sin clasificar.

Un método similar al de mínima distancia es el basado en distancia de Mahalanobis.

La distancia euclídea puede expresarse de forma matricial como

$$d_k = \sqrt{(X - \overline{X}_k)^T (X - \overline{X}_k)} \qquad (21.2.2)$$

donde X y $\overline{X}_k$ son respectivamente los vectores de valores del elemento a clasificar y de valores medios de la clase k–ésima.

La distancia de Mahalanobis es una generalización de esta, y en su forma matricial tiene la siguiente expresión:

$$d_k = \sqrt{(X - \overline{X}_k)^T C^{-1} (X - \overline{X}_k)} \qquad (21.2.3)$$

donde C es la matriz de covarianzas entre las variables estudiadas. Es decir, una matriz de la forma

$$C = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{pmatrix} \qquad (21.2.4)$$

donde σ_{ij} es la covarianza entre las variables i y j.

Máxima verosimilitud

A diferencia de los anteriores, el método de máxima verosimilitud no evalúa un parámetro geométrico dentro del espacio de atributos, sino que se basa en fundamentos estadísticos. Conociendo la media y desviación típica que caracteriza a una clase, podemos suponer una función de densidad de probabilidad, y con los valores de

un elemento dado estimar la probabilidad de que estos correspondan a cada clase. La clase con una mayor probabilidad es aquella a la que se asigna el elemento.

Este método comparte con el de distancia mínima el hecho de que todos los elementos quedan clasificados. No obstante, resulta sencillo aplicar un umbral inferior a los valores de probabilidad, de tal forma que si la clase más probable tiene un valor por debajo del umbral no se asigne el elemento a dicha clase. En este caso el método es en cierta medida semejante al de paralelepípedos, pero en lugar de estos se tienen hiperelipses n-dimensionales. Puede darse igualmente el caso de que un elemento se encuentre dentro de varias de estas hiperelipses, el cual se resuelve directamente con los valores de probabilidad de estas.

Puede verse un gráfico relativo a este método en la figura 21.4, considerando la aplicación de un umbral inferior de probabilidad.

Es importante recalcar que al emplear este método se asume que los datos tienen una distribución normal, lo cual no ha de ser necesariamente cierto según qué tipo de variables manejemos. Es importante tener en cuenta este hecho antes de utilizar este clasificador sobre nuestros datos.

21.2.2. Clasificación no supervisada

Los métodos de clasificación no supervisada no requieren del operador la definición explícita de las clases. No es necesario ningún conocimiento *a priori*, ya que es el propio algoritmo quien las define de acuerdo a los datos. Para llevar esto a cabo, es necesario suministrar algunos valores tales como el número de clases que se desea crear, los tamaños mínimos y máximos de cada una, o ciertas tolerancias para la distinción entre clases. Estos parámetros «guían» al algoritmo en la definición de clases, que se produce en estos métodos de forma simultanea a la asignación de los elementos a una u otra de dichas clases.

En general, se trata de procedimientos iterativos en los que una clasificación inicial va convergiendo hacia una final en la cual se cumplen las características buscadas de homogeneidad, número de clases, etc.

Por su propia naturaleza, estos métodos no generan clases de las cuales se conoce su significado, y será necesario estudiarlas después para saber qué representa cada una de ellas. Si en un método de clasificación supervisada definimos zonas de entrenamiento con distintas clases de suelo, el resultado sera una capa dichas con clases de suelo. Si diferenciamos según otro criterio, será ese criterio el que quede reflejado en la capa resultante. En el caso de la clasificación no supervisada, no existe tal criterio, ya que simplemente se aplican meras operaciones estadísticas con los datos, pero no se trabaja con el significado de estos. Al utilizar una zona de entrenamiento sí estamos empleando este significado, ya que le estamos diciendo al algoritmo que los valores de dicha zona representan una clase dada, esto es, que «significan» dicha clase.

Junto a la capa de clases resultantes, los métodos de clasificación no supervisada proporcionan una definición de dichas clases a través de los valores estadísticos de las mismas. Estos valores se emplearán para asignar una interpretación a cada clase una vez estas hayan sido definidas. Junto a ellas, es habitual añadir la varianza de cada clase, como indicador de la homogeneidad lograda en la clasificación.

Aunque los métodos de clasificación no supervisada son validos de por sí para establecer una separación categórica dentro de un área de estudio, es habitual que se empleen como soporte a métodos de clasificación supervisada. Mediante estos métodos se obtiene una primera división, que puede utilizarse posteriormente bien sea pa-

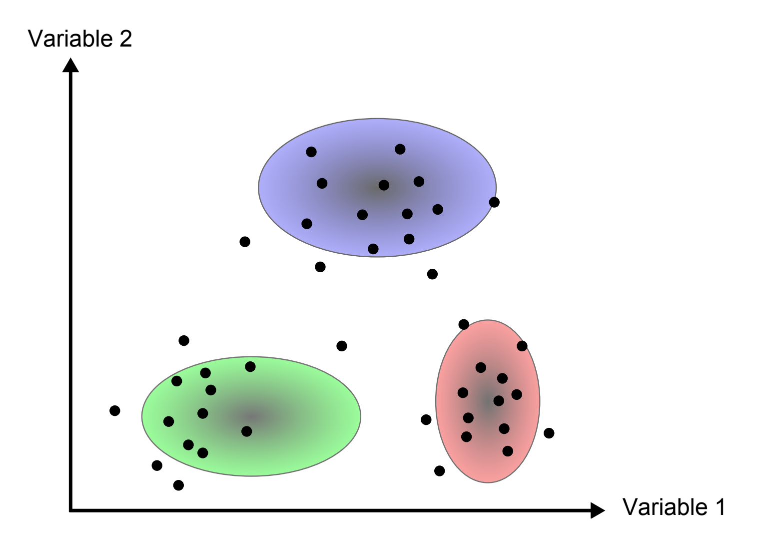


Figura 21.4: Método de clasificación por máxima verosimilitud.

ra la definición de zonas de entrenamiento o bien para operaciones más complejas como la clasificación basada en objetos que veremos más adelante.

Al igual que sucedía en el caso supervisado, existen numerosos métodos de clasificación no supervisada. La literatura estadística es rica en este tipo de formulaciones, conocidos como métodos de *clustering*[1], siendo dos de los más habituales dentro del ámbito de los Sistemas de Información geográfica los siguientes:

- Distancia mínima iterativa
- ISODATA

Distancia mínima iterativa

El método de *distancia mínima iterativa* [295] se basa en un proceso iterativo en el cual, a partir de unas clases iniciales definidas arbitrariamente, se asignan los distintos elementos a estas mediante un método de los vistos en el apartado anterior, particularmente el de distancia mínima. El número de clases iniciales es definido por el operador, y será el que aparezca en la capa resultante. Para aplicar la clasificación por distancia mínima solo es necesario conocer los valores medios de las clases, siendo este el único estadístico de cada clase con el que se va a trabajar.

Una vez que se ha realizado la asignación de clases en la primera iteración, se estudian los puntos que están incluidos en cada clase y se calculan nuevamente los valores de las medias por clase. Es decir, las clases resultantes del proceso anterior son utilizadas en cierto modo como zonas de entrenamiento. Con estos nuevos valores calculados de las medias, se vuelve a repetir la fase de asignación por distancia mínima, y así sucesivamente hasta que se cumple un criterio de convergencia.

En cada iteración del proceso hay elementos (celdas en nuestro caso, puesto que trabajamos con capas ráster) que cambian de clase. El número de dichos elementos que varían su clase se emplea como criterio de convergencia, estableciéndose un umbral inferior. Si en una iteración

[1] De *cluster* (agregado), nombre que recibe cada una de las clases o agrupaciones en las que se dividen los elementos de partida

el número de elementos que varían es menor que el umbral, se considera que el proceso ha convergido, y se detiene su ejecución. La varianza media de las clases puede utilizarse también como medida de la modificación que se produce entre el resultado de una iteración y el de la anterior. Si no se produce una disminución suficiente de la varianza, las iteraciones ya no mejoran la clasificación obtenida, con lo cual puede asumirse que el proceso ha convergido a un resultado óptimo.

El operador especifica dicho umbral de convergencia junto al número de clases deseadas, y habitualmente también un número máximo de iteraciones, con objeto de detener el proceso en caso de que no se alcance la convergencia especificada.

Este algoritmo, así como el que veremos seguidamente, clasifica todos los elementos en algún grupo, no existiendo la posibilidad de tener una capa resultante en la que existan celdas sin clasificar. De hecho, todas las celdas son utilizadas para la definición de las medias de clases sucesivas, por lo que puede decirse que todas influyen sobre la clasificación de las restantes y al final del proceso estarán asignadas siempre a alguna clase.

Como es fácil intuir, los valores iniciales de las medias de clase condicionan el proceso de clasificación, influyendo sobre la rapidez con que este converge y también sobre el resultado. Una solución habitual para establecer estos centros de clase es hacerlo equiespaciadamente en el rango que los valores a clasificar cubren dentro del espacio de atributos.

En la figura 21.5 puede verse gráficamente la forma de definir estas clases iniciales. Sobre la diagonal que atraviesa el espacio de atributos se establecen tantos puntos como clases quieran definirse. Estos serán los valores medios a los que se calculen las distancias desde cada elemento en la primera iteración.

Como puede verse, los centros de las clases no se corresponden con los centros de los grupos que existen en el conjunto de elementos, pero al cubrir el espacio y repartirse a lo largo de este, garantizan que en las sucesivas iteraciones estos centros puedan desplazarse al lugar correcto donde las clases que definan presenten una mínima variabilidad.

En la figura 21.6 pueden verse algunas etapas de un proceso de clasificación no supervisada por el método de distancia mínima iterativa, en el que se aprecian los cambios que las clases van sufriendo a lo largo de las sucesivas iteraciones.

ISODATA

El método ISODATA (Iterative Self-Organizing Data Analysis Technique) comparte los mismos fundamentos que el anterior, pero le añade algunos elementos adicionales que permiten al operador tener algo más de control sobre el proceso, al tiempo que aportan una mayor flexibilidad a los resultados.

Por una parte, puede darse el caso en que algunas de las clases establecidas no tengan suficientes elementos asignados a ellas, y no sea relevante mantenerlas. Los elementos de estas clases pueden asignarse a la siguiente clase más cercana. El método ISODATA analiza la capa resultante en busca de clases con pocos elementos, y en caso de que no superen un umbral mínimo de número de estos, los reparte entre las restantes clases. Mediante esta operación, el número de clases totales disminuye en uno.

Otro caso similar se da cuando dos de las clases resultantes son muy similares, existiendo poca distancia entre sus medias respectivas. En este caso, resulta conveniente unir dichas clases. La definición de un umbral mínimo de distancia entre clases permite al algoritmo considerar esta circunstancia.

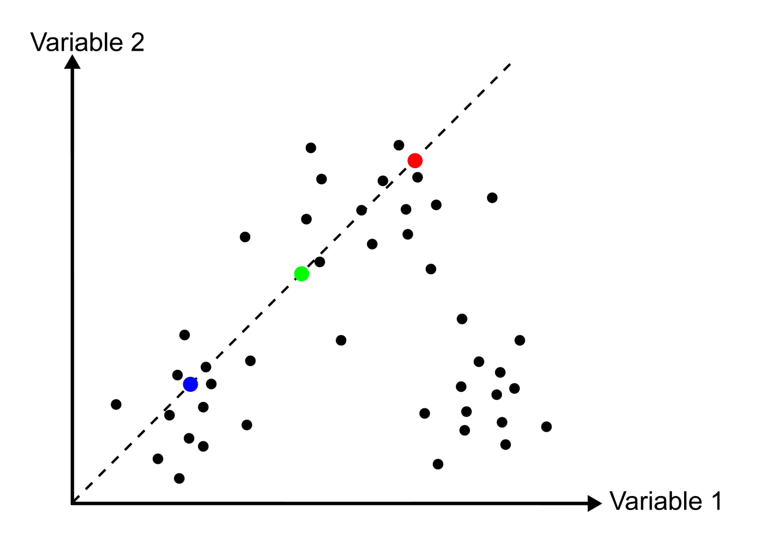


Figura 21.5: Definición de clases iniciales para el proceso de clasificación no supervisada, estableciendo los centros de clases equiespaciados en el espacio de atributos.

Un caso opuesto a los anteriores se da cuando una clase tiene una amplitud excesiva, de tal modo que la variabilidad en ella sea elevada. Puede establecerse un umbral de amplitud en función de la desviación típica, que permita localizar dichas clases, para posteriormente dividir estas en dos nuevas.

Con todo lo anterior, además de especificarse un número de clases a crear se establece un número máximo y otro mínimo de clases, para de este modo acotar el número de clases totales que resulta de los procesos de ajuste mencionados. Incluyendo estos procesos junto a la estructura del algoritmo de distancia mínima iterativa, se tiene el proceso global del método ISODATA.

Como conclusión de esta sección dedicada a la clasificación y los métodos existentes, en la figura 21.7 se muestra un ejemplo de clasificación de usos de suelo en un área de estudio, en base a imágenes de satélite y parámetros fisiográficos, llevada a cabo por algunos de los métodos descritos anteriormente.

Clustering jerárquico

Un algoritmo habitual en clasificación de objetos es la utilización de árboles jerárquicos. El proceso se basa en la creación de un árbol en el cual se disponen en sus extremos todos los objetos a clasificar, y las ramas que estos conforman se van unificando, agrupándose por similitud hasta llegar a formar una única (Figura 21.8)

El dendrograma resultante de este proceso puede utilizarse después para clasificar los elementos, simplemente descendiendo en él hasta el nivel en el que el número de clases existentes sea lo más cercano posible al deseado. Este dendrograma es más rico en información que la mera clasificación, ya que presenta distintos niveles de agrupación en lugar de uno único.

El problema con este tipo de métodos para su uso en un SIG es que los elementos individuales que clasificamos —las celdas de las capas—, se presentan en números muy elevados, del orden de millones. Manejar una estructura de este tipo correspondiente a tal cantidad de elementos es impracticable, por lo que directamente no pue-

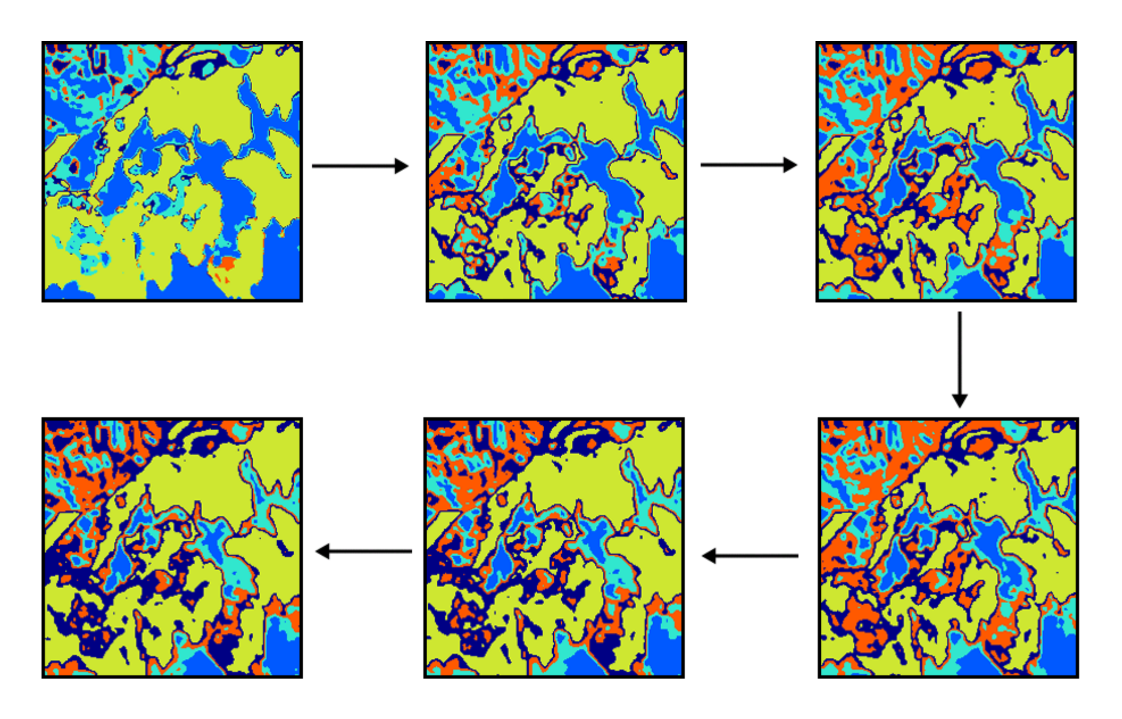


Figura 21.6: Distintas etapas de un proceso de clasificación no supervisada por distancia mínima iterativa.

de utilizarse. Es posible aplicarla, sin embargo, sobre una clasificación ya previa que reduzca el número de elementos.

Así, dadas una serie de clases, estas a su vez pueden estudiarse mediante un proceso de clustering jerárquico para disponer de información acerca de cuáles de ellas son similares y pueden unirse para simplificar la clasificación. La combinación de ambos tipos de metodologías es una práctica que permite sacar más partido a los datos de partida, analizándolos en varias etapas.

21.2.3. Incorporación del criterio espacial

Hasta este punto hemos clasificado cada elemento en función únicamente de sus propios valores. Al igual que sucede en todos los casos del análisis espacial, los distintos valores con los que trabajan tienen una referencia geográfica y existe además interacción con los valores circundantes. Cada celda de las capas ráster que clasificamos se encuentra rodeada de otras celdas, y la información de estas celdas puede ser valiosa para su clasificación debido a dicha interacción.

La información que puede utilizarse puede ser tanto la correspondiente a los datos de partida (es decir, las valores de las variables estudiadas en dichas celdas circundantes) como la resultante de la propia clasificación, ya que las clases resultantes también se presentan en un contexto espacial.

Si suponemos el sencillo caso comentado anteriormente de clasificar un área de estudio en las categorías «bosque», «cultivo» y «carretera», una celda de carretera rodeada por completo de celdas de bosque no parece lógico. Este hecho puede utilizarse como ayuda a la clasificación. Con un razonamiento similar, aquellos métodos con los que es posible que existan elementos sin

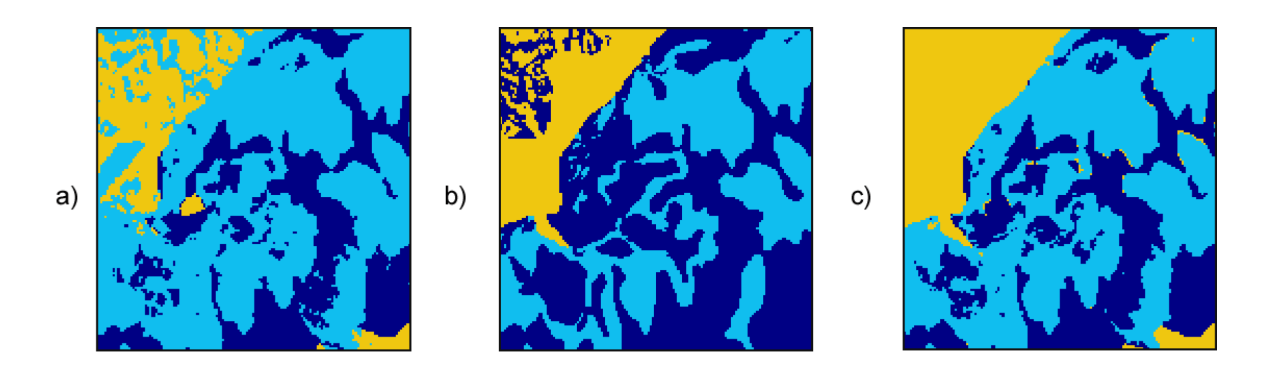


Figura 21.7: Comparación de la clasificación en clases de uso de suelo generada partir de una imagen de satélite y parámetros fisiográficos mediante los métodos de a) mínima distancia b) máxima verosimilitud y c) mínima distancia iterativa (no supervisada)

clasificar pueden tomar esta información contextual como apoyo. Una celda sin clasificar rodeada por celdas de bosque, lo más probable es que también ella pertenezca a esta tipología, a pesar de que a partir de sus valores el método de clasificación no haya sido capaz de establecer dicho resultado.

La incorporación de esta relación entre celdas vecinas puede llevarse a cabo a través de la utilización de filtros sobre la capa resultante. Un filtro de mayoría es una opción adecuada para homogeneizar la capa resultante y eliminar celdas aisladas cuya clase asignada no esté en consonancia con las de su alrededor.

La relación espacial entre las celdas que se clasifican puede incorporarse también antes de efectuar la clasificación, utilizando no solo la información puntual de las celdas sino también la *información textural*. Por «textura» entendemos la caracterización de las celdas en función de su entorno, lo cual, como puede intuirse, implica la utilización de funciones focales del álgebra de mapas. De modo similar al filtro de mayoría antes mencionado (también una función de tipo focal), la aplicación de toda la serie de filtros que vimos en el capítulo dedicado al procesado de imágenes da lugar a nuevas imágenes que pueden incorporarse al proceso de clasificación.

Para el caso de capas genéricas, no necesariamente imágenes, pueden emplearse de igual modo las distintas funciones focales tales como la media, el rango, o el valor máximo de la ventana de análisis, a seleccionar en función del significado de la variable sobre la que se aplican.

Otra forma de incorporar la relación espacial entre las celdas es la utilizada en los denominados métodos *basados en objetos*. Según estos métodos, no se clasifican las capas clasificando cada una de sus celdas, sino bloques de estas con características comunes. Estos bloques son los *objetos*, que en la fase inicial del método se estructuran de forma jerárquica a partir de esa estructura y se clasifican posteriormente.

La creación de dichos objetos a partir de las capas implica un proceso de segmentación de estas, y es ahí donde se incorpora el criterio espacial, ya que se utilizan, además de estadísticos básicos, la forma, el tamaño o los distintos descriptores de la textura de las imágenes.

Más información sobre métodos de clasificación basados en objetos dentro de un ámbito SIG puede encontrarse por ejemplo en [296] o en

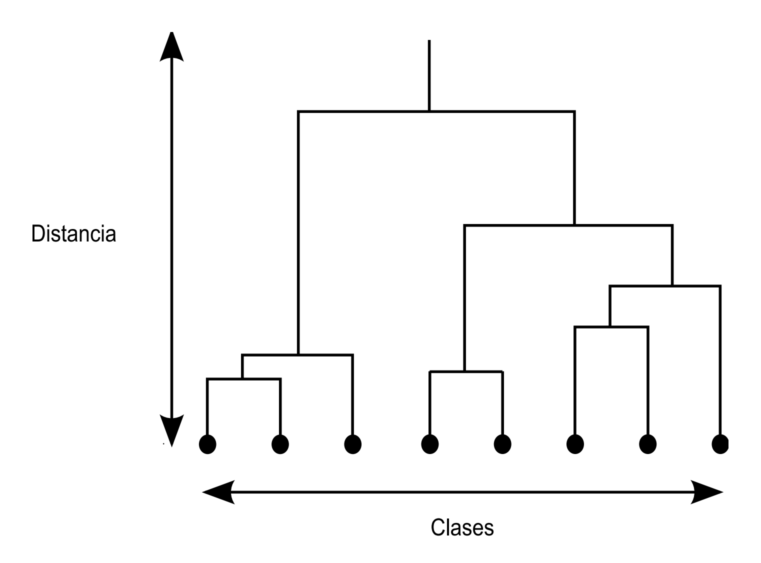


Figura 21.8: Dendrograma resultante de un proceso de clustering jerárquico

[297], este último con un tratamiento en detalle de los métodos de clasificación, así como de los de segmentación necesarios para su empleo.

21.2.4. Clasificación *débil* (*soft classification*)

En contraste con los métodos de clasificación anteriores, en particular los de clasificación supervisada, que pueden englobarse dentro de los denominados métodos de *clasificación fuerte*, encontramos otra filosofía distinta en los conocidos como métodos de *clasificación débil*. En estos, el método de clasificación no ofrece como resultado una nueva capa cuyas celdas contienen la clase a la que pertenecen, sino que la elección de dicha clase recae sobre el operador en base a los resultados que el método de clasificación produce.

Estos resultados son de forma general una serie de capas —tantas como clases a las que asignar las distintas celdas existan—, en los cuales el valor de cada celda indica la probabilidad de que dicha celda pertenezca a la clase en cuestión. De este modo, los clasificadores de tipo suave representan una etapa intermedia dentro del proceso de clasificación, ya que recogen los valores que se calculan a lo largo de este, pero no dan un veredicto posterior en base a ellos.

Si se toma el conjunto de las capas y se aplica un operador local de tipo «capa de máximo valor», el resultado que se obtendrá será una única capa, que coincidirá con la que se obtiene directamente al aplicar el método de clasificación supervisada de la forma habitual (es decir, como clasificación fuerte).

¿Cuál es la ventaja que se obtiene entonces al aplicar un método de clasificación débil y obtener las capas intermedias? En general, la información que estas capas intermedias contienen es más detallada, y además de poder emplearse para la obtención directa de la capa clasificada —el producto más habitual—, pueden utilizarse para dar una interpretación adicional a la pertenencia de cada celda a una u otra de las clases definidas.

Por ejemplo, resulta de interés conocer no solo cuál es la clase con mayor probabilidad sino también aquella que se sitúa inmediatamente después. Si clasificamos un píxel como «bosque ca-

ducifolio» y la segunda clase más probable es «bosque de coníferas», esa asignación de clase no tiene el mismo significado que si dicha segunda clase más probable es «barbecho».

De igual modo, la diferencia entre la probabilidad de la primera y la segunda clase nos dan una idea de la fiabilidad con que podemos afirmar que una celda dada pertenece a la clase a la que es asignada. En relación con este hecho se define el concepto de *incertidumbre de clasificación*, el cual para una celda dada se expresa mediante la siguiente fórmula:

$$IC = 1 - \frac{\max - \frac{S}{n}}{1 - \frac{1}{n}} \qquad (21.2.5)$$

donde max es la probabilidad máxima de entre todas las correspondientes a las distintas clases, S la suma de todas las probabilidades para la celda en cuestión, y n el numero de clases posibles.

Otra de las posibilidades que los métodos de clasificación suave aportan es la clasificación a nivel de detalle mayor que el propio píxel. El hecho de disponer de varios valores para cada píxel hace que se pueda inferir información acerca de la variabilidad que se da en el mismo, y pueden así definirse píxeles mixtos, es decir, que no pertenecen puramente a una clase, sino a varias.

Por ejemplo, si en un píxel dado tiene una probabilidad de 0,69 de pertenecer a la clase «bosque de coníferas» y un 0,31 de pertenecer a la clase «bosque caducifolio», puede entenderse que la masa boscosa del píxel esta formada por ambos tipos de especies, caducifolias y coniferas, en las proporciones que indican sus probabilidades asociadas.

Esta interpretación viene condicionada, no obstante, al cumplimiento de ciertas condiciones tales como la pureza de las zonas de entrenamiento (que no existan píxeles mixtos en los empleados para extraer las características de cada clase), circunstancia que en muchos casos es difícil de encontrar. Aun así, los valores de probabilidad de las clases, correctamente interpretados, ofrecen mayor detalle que un simple valor de clase.

21.2.5. Validación

Del mismo modo que en el capítulo 14 veíamos la manera de verificar la bondad de un método de interpolación a través del uso de puntos muestrales de comprobación, a la hora de clasificar un conjunto de capas en clases podemos contrastar el resultado obtenido si disponemos de zonas conocidas de antemano. Estas zonas se definen del mismo modo que las zonas de entrenamiento, pero no se utilizan para la clasificación, del mismo modo que no empleábamos los puntos de comprobación en el proceso de interpolación. Comparando los valores de clase de estas zonas de comprobación y los obtenidos mediante la clasificación podemos obtener índices de concordancia que permitan juzgar la corrección de dicha clasificación.

Los métodos que se emplean para esta tarea no son exclusivos de la clasificación de imágenes, sino que se adaptan de forma genérica a cualquier proceso que requiera la comparación de dos capas con información categórica. Al validar una clasificación, comparamos la capa resultante de dicha clasificación con una segunda con valores correctos de clases.

Otra circunstancia habitual en la que suelen utilizarse estos métodos es para comprobar la variación de clases a lo largo de un periodo. Comparando mediante estas técnicas la capa correspondiente al inicio del periodo con la del final de este, puede describirse y cuantificarse el cambio sufrido por las clases. Este planteamiento es muy útil, por ejemplo, para analizar el cambio en los usos de suelo de una zona.

El parámetro más sencillo que describe la similitud entre dos capas categóricas es el porcentaje de celdas coincidentes:

$$c = \frac{N_c}{N} \quad (21.2.6)$$

donde c es la concordancia global entre las dos capas, N_c el número de celdas que presentan el mismo valor en ambas capas y N el número total de celdas existentes.

Autores como [298] recomiendan al menos valores de 0.85 para considerar que la capa clasificada es operativa a efectos prácticos.

Este parámetro es sumamente sencillo y no refleja la naturaleza de la modificación que se produce entre las capas (o en el caso de utilizarse para validar una clasificación, la naturaleza del error cometido), por lo que pueden encontrarse alternativas más elaboradas.

Uno de los elementos habituales en estudio de las diferencias entre dos capas categóricas es la denominada *matriz de confusión* o *matriz de contingencias*, que ya mencionamos en el capítulo dedicado a la calidad de los datos espaciales. Se trata de una matriz de dimensiones $n \times n$, siendo n el número de clases diferentes que existen en el conjunto de las capas. El elemento (i, j) de la matriz representa el número de celdas que pertenecen a la clase i en la primera capa y sin embargo están dentro de la clase j en la segunda.

La tabla 21.1 muestra una posible matriz de contingencias para el caso de dos capas con un total de 4 clases.

Con los valores anteriores puede comprobarse entre qué clases se dan los mayores cambios (los mayores errores de clasificación) o cuáles son las que presentan una clasificación más robusta. Por ejemplo, las clases D y B parecen ser difíciles de clasificar, ya que el numero de celdas de la primera asignados a la segunda es elevado, y también al contrario. Por el contrario, la clase A parece no presentar problemas, ya que tanto su fila como su columna correspondiente presentan ambas valores bajos fuera de la celda $(1, 1)$, que es la que representa las celdas correctamente clasificados.

Este último resultado de robustez de clasificación por clases puede calcularse con la proporción de celdas correctamente clasificadas respecto a todas las clasificadas en dicha clase. La suma total de la fila k-esima dividida entre el valor de la celda (k, k) representa la anterior proporción. Es decir,

$$r = \frac{\sum_{i=1}^{n} x_i k}{x_{kk}} \quad (21.2.7)$$

Si esta misma expresión se calcula por columnas, se obtiene la proporción de celdas que, aun perteneciendo a dicha clase, han sido adjudicadas a otra distinta. Es decir,

$$r = \frac{\sum_{i=1}^{n} x_k i}{x_{kk}} \quad (21.2.8)$$

Además de estas sencillas proporciones, existen índices más complejos que pueden también calcularse a partir de los valores de la matriz de contingencias. El más habitual es el denominado *Índice Kappa* de concordancia, cuya expresión es

$$K = \frac{P_0 - P_e}{1 - P_e} \quad (21.2.9)$$

siendo P_0 la proporción total de celdas que coinciden en ambas capas, es decir,

$$P_0 = \frac{\sum_{i=1}^{n} x_{ii}}{\sum_{i=1}^{n} \sum_{j=1}^{n} x_{ij}} \quad (21.2.10)$$

y P_e calculado según la expresión

$$P_0 = \sum_{i=1}^{n} \frac{\sum_{j=1}^{n} x_{ij} \sum_{j=1}^{n} x_{ki}}{(\sum_{j=1}^{n} \sum_{k=1}^{n} x_{jk})^2} \quad (21.2.11)$$

Clase	A	B	C	D
A	20135	15	20	0
B	22	18756	133	512
C	19	70	30452	345
D	3	457	272	7018

Cuadro 21.1: Matriz de confusión.

Por su expresión, el índice Kappa no presenta sesgo por una posible coincidencia casual de clases, ya que tiene en cuenta la posibilidad de que exista concordancia por azar.

A diferencia de la proporción de celdas correctamente clasificadas, cuyo rango de valores se sitúa entre 0 y 1, el índice Kappa puede tomar valores desde -1 a 1. El valor 1 indica una concordancia completa, mientras que el -1 define una correlación de signo negativa. Valores alrededor de 0 indican que no existe correlación entre las capas. Valores por encima de 0.75 indican en general una muy buena correlación.

Al igual que para la proporción de celdas concordantes, el índice Kappa puede calcularse no solo para la capa total, sino de forma individual para cada clase.

Es de reseñar que el uso de la matriz de confusión y parámetros calculados a partir de ella como el índice Kappa no es tampoco una herramienta completa a la hora de recoger la naturaleza del error que se comete en la clasificación. Por una parte, se recoge este error en toda la extensión de la capa, mientras que puede darse de forma más notable en determinadas áreas de esta [299]. El error de clasificación no es constante a lo largo de toda la zona estudiada.

Por otra parte, la comprobación puede no ser espacialmente representativa. Si comparamos dos capas correspondientes a dos instantes distintos para ver la forma en que entre esos instantes han variado las clases presentes en la zona de estudio, utilizamos toda las celdas de la capa para la verificación. Sin embargo, al comprobar la clasificación de una serie de capas, se compara el resultado tan solo en unas zonas determinadas conocidas (si estas zonas conocidas que no empleamos en la clasificación cubrieran toda la extensión, no sería necesaria la clasificación). El hecho de utilizar zonas que conforman habitualmente bloques y son relativamente homogéneas introduce un sesgo que en general hace que la matriz de confusión presente una mayor precisión que la que realmente existe [300].

Además de las consideraciones espaciales anteriores, también deben considerarse las relaciones en el espacio de atributos. Es decir, las relaciones entre las clases. La matriz de confusión no considera estas relaciones, que sin embargo deberían tenerse en cuenta para evaluar el verdadero significado de sus valores.

Supongamos que se clasifica un área de estudio en tres clases de usos de suelo como «bosque caducifolio», «bosque de coníferas» y «lago». Si en un emplazamiento encontramos un bosque caducifolio, es un error tanto clasificarlo como bosque de coníferas como asignarlo a la clase de lago, pero no cabe duda que esta segunda posibilidad supone un mayor error. Al no existir ponderación de las celdas de la matriz de confusión a la hora de calcular índices de concordancia, este hecho no se tiene en cuenta.

El hecho de que los bosques de coníferas y caducifolios sean clases similares entre sí y los lagos sean una clase bien distinta, hace aparecer un nuevo elemento descriptor de las circunstancias que se dan en la clasificación: la *separabilidad*. El concepto es sencillo: resulta más fácil distinguir un bosque de coníferas de un lago, que hacerlo de un bosque caducifolio. Esta mayor facilidad o dificultad también se traduce a los algoritmos de clasificación como los que hemos visto.

La separabilidad mide la posibilidad de discriminar de forma efectiva entre clases distintas, y puede estudiarse mediante métodos visuales o bien numéricamente. El empleo de diagramas [301] o la utilización de parámetros como la *distancia de Jeffries–Matushita* o el *índice de Fisher* son algunas de las alternativas para evaluar la separabilidad.

Puede encontrarse más al respecto en [302], [233] o [303]

21.3. Regresión espacial

Ya hemos tratado anteriormente algunos conceptos estadísticos, y hemos comentado cómo la aplicación de estos dentro de un contexto espacial puede no ser del todo correcta, al asumirse ciertas condiciones que no se han de cumplir necesariamente (véanse los capítulos 13 y 11)

Uno de los análisis estadísticos con una presencia muy habitual en el ámbito SIG es el uso de regresiones, sean estas simples o múltiples. Como vimos en 14.2.5, a partir de los valores de una serie de predictores en un punto se puede estimar el valor en dicho punto de otra variable dada, conociendo la relación que existe entre ellas, de la forma

$$\widehat{z} = h_0 + h_1x_1 + \ldots, h_nx_n + e \qquad (21.3.1)$$

siendo $h_1, h_2 \ldots h_n$ los predictores y z la variable estimada. El parámetro e representa el error, que se supone distribuido normalmente.

Llevando a cabo este tipo de regresión, se asume que las observaciones son independientes entre sí, algo que no es en absoluto cierto en gran parte de los casos en los que se emplean datos geográficos.

El modelo anterior supone igualmente que a lo largo de la zona estudiada no existen variaciones de los parámetros estimados, es decir, que estos son constantes con independencia de la localización. Esta segunda suposición tampoco ha de ser necesariamente correcta, ya que en el contexto espacial en el que se disponen las observaciones sobre las que se basa la regresión, pueden existir variaciones locales de los parámetros de ajuste.

De existir esta variación, debe entenderse como parte del error. Adaptar las formulaciones habituales para el cálculo de regresiones al ámbito espacial en el que trabajamos requiere superar de uno u otro modo las anteriores circunstancias, y buscar la manera en que la variación no forme parte del residuo. De las soluciones existentes, una de ellas, construida sobre las anteriores ideas, es la conocida como *Geographically Weighted Regression*[2](GWR)[304]. En este modelo de regresión, la ecuación 21.3.1 se expresa de modo más genérico como

$$\begin{aligned} \widehat{z}(u,v) &= h_0(u,v) + h_1(u,v)x_1 \\ &\quad + \ldots, h_n(u,v)x_n \\ &\quad + e(u,v) \end{aligned} \qquad (21.3.2)$$

En este caso, también los parámetros estimados dependen la localización, que viene expresada a través de las coordenadas u y v

La estimación de estos parámetros exige también adaptar el método de Mínimos Cuadrados

[2]Regresión Ponderada Geográficamente

Ordinarios, utilizado habitualmente para la estimar los de la ecuación 21.3.1. Los parámetros para un modelo de regresión lineal se obtienen según la expresión matricial

$$h = (X^TX)^{-1}X^TY \quad (21.3.3)$$

Añadiendo una ponderación que dependa a su vez de la localización, tenemos la siguiente expresión, que permite calcular los parámetros de la ecuación 21.3.2.

$$h = (X^TW(u,v)X)^{-1}X^TW(u,v)Y \quad (21.3.4)$$

siendo W los pesos a utilizar.

Estos pesos se toman de tal forma que las observaciones situadas más cerca del punto donde desean estimarse los parámetros tienen mayor influencia. Este tipo de formulaciones de ponderación en función de la distancia ya las hemos visto en el capítulo 14, tanto para la interpolación como para el cálculo de densidades. En el caso del GWR, una función habitual es la siguiente:

$$w_i(u,v) = e^{\left(\frac{-d}{h}\right)^2} \quad (21.3.5)$$

donde d es la distancia entre las coordenadas de la observación y (u,v), y h es la *anchura*. Este parámetro es el equivalente al radio máximo de influencia que veíamos para el cálculo de densidad empleando un núcleo gaussiano.

21.4. Evaluación multicriterio y combinación de capas

La combinación de capas es una operación muy habitual. Diferentes variables, cada una de ellas recogida en una capa, se combinan para obtener algún tipo de resultado en base a la información que representan. Dentro de un SIG, conocemos ya operaciones de combinación de capas tanto en formato ráster (mediante el álgebra de mapas y sus funciones focales) como en formato vectorial (mediante operaciones de solape). Es, no obstante, con las primeras con las que podemos plantear expresiones complejas que incorporen esas variables, tal como, por ejemplo, la Ecuación Universal de Pérdidas de Suelo (USLE) que vimos en su momento en el apartado 15.1.

En algunos casos, como el de la USLE, partimos de una formula definida en la que no cabe modificación alguna (este es el caso habitual con fórmulas empíricas o de base física). La formula contiene una serie de variables y cada una de estas se encuentra recogida en una capa. Basta aplicar dicha fórmula mediante una operación local del álgebra de mapas, y el resultado es una capa con la variable resultante de dicha fórmula.

En otras ocasiones, la fórmula no viene definida de antemano, sino que disponemos de una serie de variables que tienen influencia sobre un determinado fenómeno y necesitamos combinarlas para obtener una nueva variable que nos aporte información sobre ese fenómeno. El objetivo es agregar las variables de las que disponemos, como factores implicados en el proceso que analizamos.

Esta última situación es muy habitual dentro de un SIG, en los denominados procesos de *evaluación multicriterio*. En estos, se dispone de una serie de parámetros que afectan a un determinado fenómeno, y se pretende evaluar la medida en que afectan y condicionan al mismo, para de este modo estudiarlo y, en la mayoría de los casos, tomar decisiones en función de los resultados obtenidos. Cada una de las variables que influyen se recoge de forma habitual como una variable en una capa independiente, y se debe en primer lugar juzgar cómo los valores de cada variable afectan al fenómeno, y después combinar todas esas afecciones en una única variable que sea la que ayude en la toma de decisiones.

La evaluación multicriterio es la base de, por ejemplo, el análisis de idoneidad, un análisis muy frecuente en el ámbito SIG, y en el cual se pretende localizar dentro de una zona de estudio los mejores emplazamientos para una determinada actividad. Cada uno de los factores que influyen en el desarrollo de dicha actividad son valorados de forma independiente, y después conjugados mediante una expresión matemática.

Estas formulaciones incorporan elementos probabilísticos diversos, y existe un gran número de metodologías para formular los modelos que nos llevan a evaluar la idoneidad de cada localización. Será en estos elementos en los que nos centremos en este apartado del capítulo, con objeto de extender la potencialidad del álgebra de mapas para la realización de este tipo de operaciones. Aunque una aplicación fundamental de todos ellos es la evaluación multicriterio (y por ello están incorporados en este apartado), tienen utilidad también en otro tipo de modelos, tales como, por ejemplo, modelos de predicción de cambios en el uso de suelo, también frecuentes en el campo de los SIG.

En líneas generales, presentaremos formulaciones que nos permitan combinar las variables de forma más elaborada, para poder crear modelos geográficos de cualquier índole (como por ejemplo los modelos de idoneidad mencionados) y más precisos que los que pueden obtenerse con los elementos que hemos visto hasta el momento.

Dos son los apartados en los que ampliaremos nuestros conocimientos sobre la elaboración de modelos como los anteriores:

- La creación y preparación de las capas que reflejan los distintos criterios a aplicar
- La forma de combinar esos criterios en una expresión matemática.

21.4.1. Creación de capas a combinar

A la hora de plantear un modelo con diferentes variables y criterios, necesitamos expresar de forma numérica el valor de esos criterios que posteriormente agregaremos. En el caso de un modelo de idoneidad, por ejemplo, necesitamos crear una capa que nos diga si la actividad puede o no llevarse a cabo en función de cada criterio.

Como vimos en el capitulo dedicado al álgebra de mapas, además de las operaciones aritméticas podemos aplicar de igual modo operaciones lógicas a la hora de combinar varias capas. Este tipo de conceptos lógicos también estaban implícitos en las operaciones de solape entre capas vectoriales, según también vimos en su momento. Nos servían, por ejemplo, para eliminar de una capa todas las zonas a menos distancia de un cauce que el Dominio Publico Hidráulico (DPH) de este, de tal modo que restringíamos las zonas donde podíamos establecer una edificación a aquellos puntos fuera de dicho DPH. Esa distancia la calculábamos realizando un área de influencia, proceso que podríamos de igual modo llevar a cabo con capas ráster.

Ese era un modelo de idoneidad muy sencillo, con un único criterio: la distancia al cauce. Esta distancia daba lugar a dos posibles estados: o bien un punto está dentro del DPH (no se puede construir en él), o bien está fuera (se puede construir en él). Las operaciones lógicas nos sirven para expresar esto, y mediante ellas podemos desarrollar nuestro modelo.

No obstante, y sin necesidad de añadir más criterios que compliquen el modelo (es decir, sin necesidad de que la evaluación sea multicriterio, sino por el momento monocriterio), podemos encontrar situaciones en las que la lógica booleana no refleja con suficiente precisión un criterio dado. Sigamos utilizando el criterio de distancia, pero en este caso supongamos el siguiente caso: buscamos un lugar donde emplazar una fabrica

y conocemos el emplazamiento del principal núcleo urbano. En este caso debemos igualmente mantenernos alejados de la ciudad para evitar las afecciones que la fábrica puede causar sobre ella. Por otra parte, sin embargo, no interesa situarla a demasiada distancia, ya que entonces será muy costoso acceder a los servicios de las ciudades.

Igual que en el caso anterior, podemos definir una distancia fija por debajo de la cual no debemos construir, y añadir además una distancia límite por encima de la cual tampoco resulta rentable económicamente hacerlo. De esta forma, planteamos nuestro sencillo modelo aunque, como veremos, es fácilmente mejorable.

Supongamos que establecemos esa distancia mínima en 4000 metros y la máxima en 15000. Un punto situado a 4001 metros es perfectamente viable, mientras que uno a 3999 no lo es. De igual modo, ese punto situado a 3999 metros es igual de inviable que uno situado a tan solo 2 metros, y puntos a 4001 metros y 14999 metros son igualmente viables, sin que exista distinción entre ambos. Estos casos no parecen muy lógicos a primera vista, y esto es debido a que, en realidad, no existen solo dos posibles clases como estamos planteando en el modelo. Un punto puede ser no solo *adecuado* o *no adecuado*, sino que existe toda una gama de posibles categorías tales como *muy adecuado*, *poco adecuado*, *completamente inviable* o *adecuado pero casi inviable*.

En el caso del DPH, este criterio plantea una *restricción*. A efectos de este criterio no importa el valor de la distancia, sino tan solo si es mayor o menor que el umbral, que es lo que, en base a la legalidad vigente, nos permitirá construir o no. En el ejemplo que ahora proponemos, sin embargo, no tenemos una restricción, sino una variable que condiciona, y este condicionamiento puede darse a diferentes niveles, presentando más de dos alternativas posibles.

Un caso similar lo podemos encontrar si tenemos un MDE y conocemos la altitud a la que aparece una determinada especie de planta. Esta altitud vendrá definida por un limite inferior y uno superior, pero ello no quiere decir, lógicamente, que un metro por debajo de la cota inferior o un metro por encima de la superior sea imposible encontrar dicha especie, ni que dentro de ese rango óptimo sea igual de probable encontrarla con independencia de si nos situamos cerca o no de los límites establecidos. Para reflejar este hecho necesitamos, igual que en el caso anterior, poder expresar que no solo existen zonas *óptimas* o *inviables* para una especie, sino que pueden existir otras zonas *adecuadas* o zonas *poco adecuadas pero no completamente inviables*.

Como vemos, el lenguaje natural nos ofrece una variedad de posibilidades que la lógica booleana, con su binomio verdadero/falso no nos ofrece. El objetivo es, pues, trasladar toda esa gama de posibilidades a una forma compatible con el manejo de capas dentro de un SIG y con la realización de operaciones entre conjuntos, del mismo modo en que lo hacíamos al aplicar el álgebra booleana. La solución a esto es la denominada *lógica difusa*.

La lógica difusa pretende acomodar las ideas anteriores al ámbito matemático, y que las operaciones lógicas no solo se basen en *sí* (verdadero) y *no* (falso), sino también en opciones intermedias (*quizás*)[3] Esto nos va a permitir desarrollar modelos más precisos, y nos dará la posibilidad de aplicar las herramientas algebraicas que ya conocemos, más potentes en este sentido que las

[3]Esta es una definición muy informal de la lógica difusa, la cual puede definirse con todo el rigor matemático necesario. No obstante, un enfoque más riguroso excede las necesidades de este apartado, y aquí veremos tan solo la aplicación más práctica de las ideas básicas, orientando estas hacia su utilización dentro de un SIG. Información mucho más extensa y definiciones rigurosas sobre lógica difusa pueden encontrarse en [305, 306].

operaciones lógicas. Además, podremos aplicar los elementos que veremos en la siguiente sección, para combinar las capas que gracias a la lógica difusa obtendremos.

Básicamente, la idea es que, si antes expresábamos esa naturaleza verdadera o falsa de un elemento con total probabilidad (cuando era *verdadero* existía una probabilidad del 100 % de que la planta se diera a esa altitud, y cuando era falso esa probabilidad era del 0 %), ahora queremos que los valores de probabilidad no sean un conjunto finito de dos elementos, sino que puedan situarse en todo el rango de valores posibles.

La idea de esa variable «difusa» que queremos obtener es similar a la de probabilidad, y ambas se expresan como un valor entre 0 y 1, aunque conceptualmente presenten diferencias. La probabilidad nos expresa en qué grado resulta factible que se produzca un fenómeno (por ejemplo, qué probabilidad hay de que aparezca una especie de planta en función de la altitud), mientras que la variable difusa nos expresa una incertidumbre acerca de la pertenencia de un elemento a una clase (por ejemplo, en qué medida un punto, en función de su distancia a una ciudad, puede incluirse dentro de los puntos viables para establecer una fábrica). Este tipo de variables resultan, por tanto, de utilidad siempre que las clases con las que trabajamos no tengan una frontera bien definida, sino que exista una transición continua entre ellas.

Este análisis resulta similar en cierta medida a lo que veíamos en el apartado dedicado a los clasificadores suaves, donde teníamos distintas capas que nos indicaban la pertenencia de un elemento a cada una de las clases definidas. La clasificación suave nos muestra los pasos intermedios de un proceso completo, el de clasificación, en el cual a partir de las variables de partida, se obtiene una capa resultante.

Como indica [307], existen dos formas de plantear estos modelos: por un lado, tratar las variables por separado y después agregarlas, o bien establecer un planteamiento holístico que trate todas estas variables como un conjunto. La clasificación está en este último grupo. Las operaciones que vemos dentro de este apartado se encuentran, sin embargo, dentro del primero.

Un elemento clave en la lógica difusa son las funciones que nos permiten calcular los valores de la variables difusa (es decir, aplicar el criterio concreto sobre cada variable). Puesto que el valor de esta nos indica la certidumbre con la que podemos afirmar que un elemento es miembro de una clase dada, las anteriores funciones se conocen como *funciones de miembro*. A partir de un valor dado asignan uno nuevo entre 0 (se sabe con certeza que el elemento no pertenece a la clase) y 1 (se sabe con certeza que sí pertenece a la clase).

Por ejemplo, para el supuesto de utilizar la distancia a una ciudad como variable base, una posible función de miembro es la mostrada en la figura 21.9.

Para definir esta función de miembro, se han establecido, además de los valores límite, un rango de valores óptimos (entre 6000 y 10000 metros de distancia), que son los que delimitan las zonas idóneas para situar nuestra fábrica. En este rango, se tiene un valor 1. Desde los valores óptimos, los valores descienden, haciéndolo de tal modo que toman un valor igual a 0,5 en el punto en el que situábamos los valores límite. Esto es lógico si pensamos que hay que definir un punto umbral a partir del cual considerar si el elemento está dentro o fuera de la clase (para así tener la clasificación dicotómica de verdadero y falso), y ese punto resulta natural establecerlo en la mitad del intervalo.

Junto con la gráfica de la función de miembro anterior, se muestra la forma que tendría una

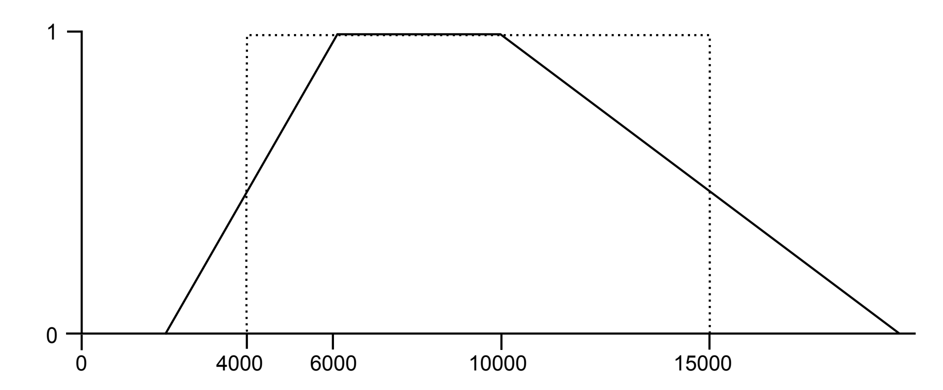


Figura 21.9: Un ejemplo de función de miembro. En trazo punteado, aspecto de una función de miembro que asignara tan solo valores *verdadero* y *falso*

.

función de miembro[4] que se comportase igual que si operáramos con los elementos del álgebra booleana, restringiendo los valores posibles a dos: *verdadero* y *falso*. Esta función presenta un salto brusco del 0 al 1, de tal modo que no es posible asignar ninguno de los valores intermedios. Para cualquier valor dado, el nuevo valor que se obtiene al aplicar esta función es, o bien 0, o bien 1. El salto de la función se produce exactamente en los valores límite, justamente donde la verdadera función de miembro toma el valor 0,5.

La diferencia entre los resultados que se obtienen al aplicar una función de miembro como la anterior y aplicando una mera clasificación en dos clases de distancia pueden apreciarse en las capas de la figura 21.10. Junto con una capa de distancia a un punto dado, se muestran las resultantes de, en base a dicha capa, aplicar un criterio en forma de restricción con dos clases posible (zonas dentro del intervalo óptimo *vs.* zonas fuera del intervalo) o bien creando una variable que refleje la certidumbre de pertenencia a cada una de las clases anteriores.

La función de miembro puede ser cualquier función, y no necesariamente similar a la que se muestra en la figura 21.9. Basta con que cumpla las siguiente condiciones:

- Estar acotada entre 0 y 1
- Asignar valor 1 a los elementos que indudablemente pertenecen al conjunto o clase
- Presentar un descenso «suave» desde los elementos con valor 1 hasta los restantes.

Es habitual, no obstante, encontrarnos con situaciones como las anteriores, en las que tengamos dos rangos, uno para el óptimo dentro del cual tendremos valor 1 (sea este rango acotado por los valores b y c, $b < c$), y otro más amplio que abarca todos los valores distintos de 0 (entre los valores a y d, $a < d$). En esta situación, lo único necesario es definir las transiciones desde el óptimo hasta los límites exteriores, es decir, los descensos suaves anteriormente citados.

En el ejemplo de la figura esta transición es lineal, y la función de miembro se puede definir de la siguiente manera:

[4]Realmente no es una función de miembro, ya que no cumple las características de una de ellas, pero la denominaremos así para facilitar la comprensión de esta parte.

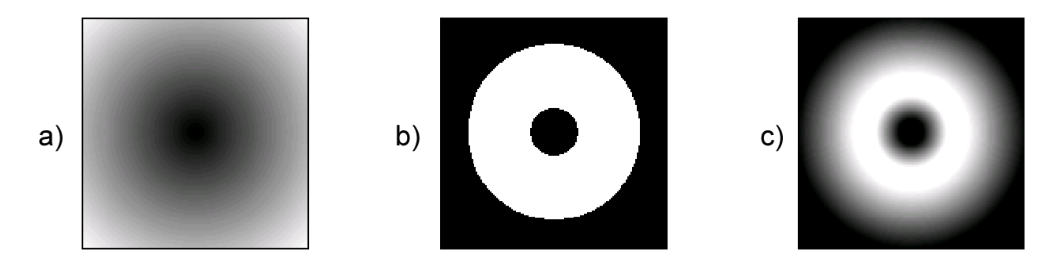


Figura 21.10: a) Capa de distancias a un punto. b) Separación en dos clases de viabilidad en función de la pertenencia a un intervalo óptimo de distancia (en blanco zonas viables, en negro zonas inviables). c) Certidumbre de pertenencia a la clase viable aplicando la función de miembro de la figura 21.9

.

$$f(x) = \begin{cases} 0 & \text{si } x < a \\ \frac{x-a}{b-a} & \text{si } a \leq x \leq b \\ 1 & \text{si } b < x < c \\ \frac{d-x}{d-c} & \text{si } c \leq x \leq d \\ 0 & \text{si } x > d \end{cases} \qquad (21.4.1)$$

Pueden elegirse cualesquiera valores para los parámetros a, b, c y d, obteniéndose toda una familia de curvas distintas. Por ejemplo, el intervalo óptimo puede reducirse a un único punto ($b = c$), en cuyo caso tendríamos una función triangular.

Otra solución habitual es emplear una función sigmoidal para las transiciones, quedando la definición global de la función de miembro como sigue:

$$f(x) = \begin{cases} 0 & \text{si } x < a \\ \frac{1}{2}\left(1 + \cos\left(\pi \frac{x-a}{b-a}\right)\right) & \text{si } a \leq x \leq b \\ 1 & \text{si } b < x < c \\ \frac{1}{2}\left(1 + \cos\left(\pi \frac{d-x}{d-c}\right)\right) & \text{si } c \leq x \leq d \\ 0 & \text{si } x > d \end{cases}$$

Una función de distribución de probabilidad gaussiana (esto es, una campana de Gauss), también puede utilizarse como función de miembro. En este caso, el óptimo sería el punto que coincide con la media de dicha distribución.

21.4.2. Métodos de combinación de capas

A la hora de plantear un modelo, utilizaremos habitualmente no una única capa como en los ejemplos anteriores, sino un conjunto de ellas, pues serán con seguridad varios los factores que influyen en el fenómeno que estudiamos. Si pretendemos buscar el emplazamiento de una fábrica, la distancia a la ciudad más próxima condiciona la idoneidad de cada localización, pero también lo harán el tipo de suelo, el uso de suelo, la pendiente, y otra serie de factores.

De igual modo, la probabilidad de que una especie aparezca en un determinado punto depende de la altitud, pero también de la insolación o las características del suelo, entre otros factores. Cada uno de ellos se pueden tratar por separado, y obtenerse capas como las que hemos obtenido en el apartado anterior, con valores entre 0 y 1, que indiquen un valor de probabilidad o bien la pertenencia a un conjunto difuso. Con todas estas capas es con las que debemos plantear el modelo, agregando la información que nos suministran en una nueva variable.

Las operaciones algebraicas más sencillas nos dan una primera herramienta para esa agregación. Estadísticos como la media aritmética, la media armónica, o los valores máximo o míni-

mos del conjunto pueden servir para combinar en un único valor los valores de los distintos criterios [308].

La mera suma de los distintos valores es empleada también con frecuencia. Si los valores representan probabilidades, esta suma corresponde al operador O. Si en lugar de la suma empleamos el producto, esto es equivalente a la operación Y, es decir, la intersección de los conjuntos (estas operaciones tienen el mismo significado si las aplicamos sobre elementos verdadero/falso codificados con valores 0/1, es decir, si la combinación es en esencia booleana).

Una opción más elaborada es una media ponderada de los distintos factores, de la forma

$$y = \frac{\sum_{i=n} N a_i x_i}{\sum_{i=n} N a_i} \qquad (21.4.2)$$

Esta es una solución habitual en la evaluación multicriterio [307].

Además de estos métodos, existen otros que, según las circunstancias, pueden resultar más adecuados para una correcta combinación de los factores considerados. Entre ellos, encontramos la denominada *regla de Dempster*. En el caso de aplicar una media ponderada como la de la ecuación 21.4.2, la elección de los pesos a_i no es trivial, especialmente cuando el número de factores es elevado. La metodología de *Jerarquías Analíticas*, que también veremos, nos ayudará a establecer dichos pesos de forma coherente con la importancia de cada factor.

Regla de Dempster

La regla de Dempster esta basada en la denominada *Teoría de la Evidencia*, una extensión de la teoría bayesiana de la probabilidad [309].

El elemento básico para agregar las distintas evidencias dentro de esta teoría es la regla de Dempster. La característica principal de esta regla es que, al contrario que el producto de las probabilidades, no disminuye a medida que se agregan más y más factores (en el producto, por ser la probabilidad siempre menor que 1, sí sucede así). En particular, si se agregan dos valores mayores que 0,5 el resultado es mayor que ambos. Si se agregan dos valores menores que 0,5 el resultado es menor que ambos. Si uno es mayor y otro menor, el resultado es intermedio entre ambos.

Las probabilidades se denotan con la letra m (de *masa*) y se conocen como *Asignación de Probabilidad Básica*.

El lector interesado en profundizar en el tema puede encontrar en [309, 310] exposiciones más detalladas al respecto. Más ejemplos de la regla de Dempster y de numerosas variantes de la misma pueden encontrarse en [311].

Jerarquías Analíticas

Una suma ponderada de los distintos factores es uno de los métodos más habituales de combinar estos. Si los factores son solo dos, o incluso tres, no resulta difícil asignar los pesos a cada uno de ellos conociendo la importancia relativa que tienen. Sin embargo, cuando son más numerosos (lo cual sucede muy habitualmente), asignar pesos de forma consistente no resulta sencillo, ya que las relaciones uno a uno entre los distintos factores son demasiadas como para poder tener una visión global de ellas. Es necesario para ello recurrir a alguna metodología con cierta sistematicidad.

La más popular de estas metodologías es la de las denominadas *Jerarquías Analíticas*[5] [312]. En esta metodología, no han de asignarse pesos directamente a todos los factores, sino tan solo hacerlo para cada una de las posibles combina-

[5]En inglés, *Analytical Hierarchy Process (AHP)*

ciones entre dos de dichos factores. Puesto que se reduce la ponderación de todos los factores simultáneamente a una serie de sencillas ponderaciones entre dos elementos, resulta mucho más sencillo establecer la importancia relativa de cada factor, a la par que mucho más preciso.

Los pesos de las distintas parejas de factores se colocan en una matriz de dimensiones $n \times n$, siendo n el número de factores que se quieren ponderar. El valor en la posición (i, j) representa la ponderación entre el factor i y el factor j, esto es, la importancia relativa de este primero frente al segundo. En esta matriz (sea M), se cumple siempre que $M_{ij} = \frac{1}{M_{ji}}$, y $M_{ii} = 1$.

Existen estudios psicológicos que muestran que no se puede comparar simultáneamente más de 7 ± 2 elementos, y en base a este hecho los autores de esta metodología recomiendan utilizar valores entre 1 y 9, según lo mostrado en el cuadro 21.4.2.

Una vez creada la matriz de comparaciones, el autovector principal de la misma contiene los distintos pesos a asignar a cada uno de los factores.

Veamos un ejemplo. Supongamos tres capas con tres factores, sean a, b y c, y la siguiente matriz de comparaciones entre pares de ellos, que describe la importancia de cada uno de los mismos frente a los restantes.

	a	b	c
a	1	4	5
b	1/4	1	1/2
c	1/5	2	1

Según la tabla, el factor a es cinco veces más relevante que el c y, consecuentemente, c tiene un quinto de la importancia de a.

Factores y restricciones

A la hora de combinar una serie de capas, el papel que estas juegan en el modelo puede ser bien distinto en función de la variable que contengan. Por ejemplo, existiendo una limitación legal a la construcción cerca de un cauce, trabajar con una capa de distancia al cauce con las herramientas que acabamos de ver para preparar y combinar capas no aporta ninguna ventaja adicional. La distancia no es en este caso un factor cuya influencia pueda graduarse, sino una restricción que simplemente nos servirá para saber si es posible o no construir en un emplazamiento dado.

En el caso de las restricciones, las operaciones lógicas con las que trabajábamos son suficientes para excluir aquellas zonas en las cuales no va a ser viable desarrollar una actividad, con independencia del resto de variables. En el ejemplo de la distancia al cauce para establecer una edificación, no importa que el resto de factores sean apropiados y el lugar sea idóneo para edificar. Si la distancia al cauce es menor que la establecida por el Dominio Público Hidráulico, no resulta viable como emplazamiento.

Frente a aquellos parámetros que representan restricciones, existen otros que afectan a la actividad que evaluamos en el modelo, pero su influencia puede tratarse como una variable continua, tal y como hemos visto. La combinación de estos mediante operaciones lógicas no es suficiente para reflejar este hecho, y deben por ello aplicarse operaciones aritméticas o algunas de las metodologías alternativas que hemos visto.

Esta distinción entre factores y restricciones es importante para una correcta integración de todas las capas con las que vamos a trabajar, dándole a cada una el papel que le corresponde en el modelo. Se pueden tratar las capas de factores por separado y después aplicarles las restricciones, conjugando así los dos tipos de parámetros que se consideran en el modelo.

Las operaciones lógicas pueden evaluarse como operaciones aritméticas, aplicando el opera-

Valor	Descripción
1	Misma importancia
3	Predominancia moderada de un factor sobre otro
5	Predominancia fuerte
7	Predominancia muy fuerte
9	Predominancia extrema
2, 4, 6, 8	valores intermedios
Valores recíprocos	Valores para comparación inversa

Cuadro 21.2: Valores de comparación en la metodología de jerarquías analíticas

dor multiplicación y codificando con valores 0 y 1 los valores lógicos *verdadero* y *falso*. El ejemplo de la figura 21.11 muestra un supuesto de combinación de dos factores y una restricción en un modelo sencillo.

21.5. Análisis de Componentes Principales

El Análisis de Componentes Principales (ACP, también conocido como *transformación de Kahunen-Loeve o de Hotelling* es otra técnica estadística con gran importancia dentro de los SIG, en especial, y de modo similar a la clasificación (aunque al igual que entonces, no exclusivamente), en el trabajo con imágenes.

El ACP es una transformación que pretende disminuir la dimensionalidad de un conjunto de variables, reduciendo este a uno más pequeño de forma que se pierda la menor información posible. Se trata de «resumir» la información que esas variables contienen, pero eliminando partes redundantes como por ejemplo las derivadas de la dependencia que pueda existir entre las variables. Se busca, pues, eliminar datos sin eliminar información.

Esta trasformación es útil ya que disminuye el volumen de datos total, facilitando las operaciones de análisis e interpretación de las variables, así como su propio manejo.

En el ámbito del SIG, trabajamos con frecuencia con muchas capas, y una reducción en el número de estas facilita el planteamiento de modelos tales como los que veíamos en el apartado anterior. No solo disminuyen los cálculos a efectuar y la complejidad de las operaciones, sino que resulta más sencillo interpretar las relaciones entre variables cuando estas no se presentan en gran número (recuérdese, por ejemplo, lo que se comentó en el apartado 21.4.2 en relación a la metodología de Jerarquías Analíticas)

En el caso de imágenes con elevado número de bandas (multiespectrales e hiperespectrales), vimos en el capítulo 17 que la representación de estas no puede hacerse empleando todas las bandas, sino que deben prepararse composiciones con, a lo sumo, tres bandas. Reducir un conjunto de muchas bandas a uno de tres con la máxima información posible, de forma que ya pueda ser representado y analizado visualmente, resulta una metodología más adecuada que elegir tres bandas cualesquiera de ese mismo conjunto. Por su propia naturaleza, es probable además que esas bandas estén altamente correlacionadas y contengan información redundante, o al menos con cierta redundancia de cara al trabajo concre-

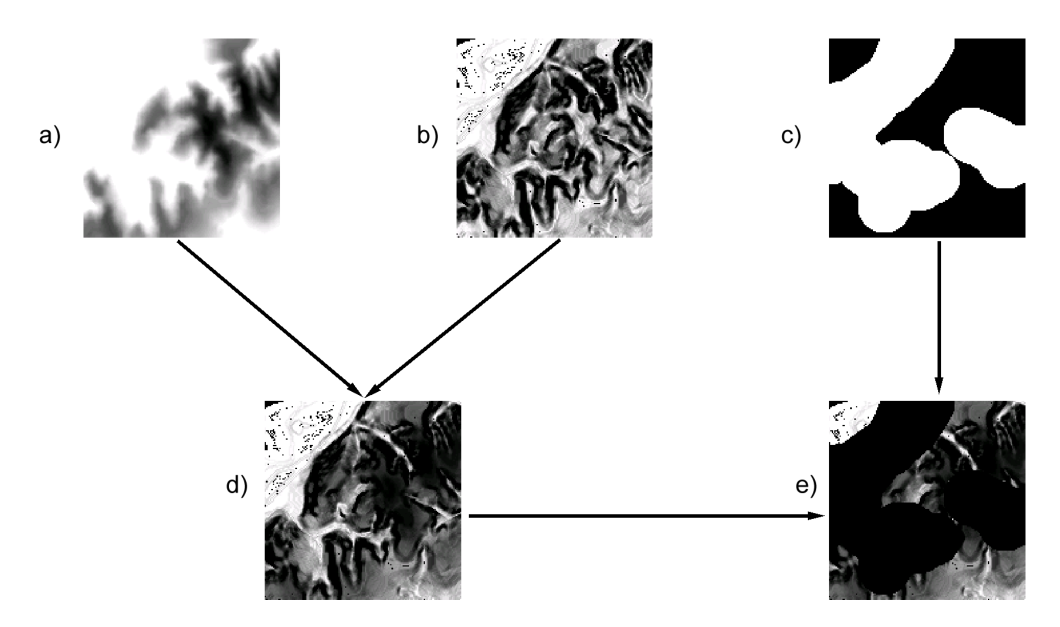


Figura 21.11: Combinación de factores y restricciones en un modelo sencillo. Los factores a) y b) se combinan para obtener un resultado ponderado d). A este se le aplica después la restricción c) para obtener el mapa final e).

.

to que vayamos a llevar a cabo con la imagen en cuestión.

Por las razones anteriores, las imágenes son un tipo de dato que se presta en gran medida a la aplicación de este tipo de transformación, aunque cualquier conjunto de capas puede trasformarse de manera similar.

El planteamiento conceptual de la transformación puede resumirse de la forma siguiente. Sea un conjunto de n variables. Estas definen un espacio vectorial n-dimensional, de tal forma que las características de un elemento dado (en caso de una capa ráster será una celda dada) se expresan mediante un vector de n elementos de la forma $(x_1, x_2, \ldots, x_n)$. El ACP busca definir un cambio de base en ese espacio n−dimensional, de modo que los vectores de la nueva base guarden una relación directa con las direcciones de variabilidad del conjunto de datos. El primer vector de la base señala la dirección de máxima variabilidad, el segundo la segunda dirección de máxima variabilidad, y así sucesivamente. Eso quiere decir que, en esta nueva base, la mayor parte de la información se va a encontrar en la dirección del primer vector, y que esta cantidad de información va a decrecer paulatinamente según tomamos cada uno de los sucesivos vectores de la base.

Al aplicar el cambio de base a un vector $(x_1, x_2, \ldots, x_n)$, se obtiene un nuevo vector $(x'_1, x'_2, \ldots, x'_n)$, expresado en las coordenadas de la nueva base. Puesto que las primeras coordenadas de este nuevo vector se corresponden con las direcciones de máxima variabilidad, podemos tomar solo las primeras p coordenadas (siendo $p < n$) y tener un vector de la forma $(x_1, x_2, \ldots, x_p)$, sabiendo que haciendo esto estamos perdiendo poca información a pesar de reducir la dimensión del vector original. Estas p coordenadas son los p componentes principales.

Es importante reseñar que las variables resultantes no tienen significado físico alguno aun en el caso de que las variables originales sí lo tuvieran, no pudiendo utilizarse en ese sentido del mismo modo que dichas variables originales.

Aunque resulta de interés tomar las componentes principales y descartar los vectores finales de la base, estos también pueden aportar información relevante según qué análisis se quiera realizar. La información sobre los ejes principales define aquella información que aparece en todas las variables utilizadas. La información restante es la que aparece solo en alguna de ellas. Esto puede utilizarse, por ejemplo, para estudiar los cambios producidos a lo largo del tiempo.

Si tomamos como n variables un conjunto de n medidas de una única variable en n instantes, el análisis de componentes nos permite separar la información común a todos ellos de aquella que solo corresponde a algunos. De este modo, podemos analizar una u otra parte de la información según sean los resultados que busquemos.

En [313] y [314] pueden encontrarse dos ejemplos del uso de transformaciones de componentes principales en la detección de cambios de uso de suelo a partir de series de imágenes.

Para calcular la matriz del cambio de base que define la transformación, se puede partir de la matriz de covarianzas C (ver ecuación 21.2.4) o bien de la matriz de correlación ρ, en la cual los elementos son los coeficientes de correlación de Pearson. Es decir,

$$\rho = \begin{pmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & \rho_{22} & \cdots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \cdots & \rho_{nn} \end{pmatrix} \qquad (21.5.1)$$

donde ρ_{ij} se calcula según

$$\rho_{ij} = \frac{\sigma_{ij}}{\sigma_{ii}\sigma_{jj}} \qquad (21.5.2)$$

Esta ultima matriz se emplea en lugar de la de covarianzas en caso de que las unidades en que se miden las variables no guarden relación entre sí, con lo cual no es posible compararlas.

El siguiente paso una vez que se tiene la matriz es la obtención de sus autovalores y autovectores. Estos autovectores son los vectores de la nueva base, y sus autovalores asociados nos sirven para establecer el orden en que han de considerarse. Así, el mayor autovalor indica que su vector propio asociado es aquel cuya dirección es la de la máxima variabilidad, y el de mínimo valor se asocia, de modo similar, al vector en la dirección de mínima variabilidad.

Para más detalles, [44] ofrece información detallada sobre el análisis de componentes principales y su uso en el campo de la teledetección.

21.6. Resumen

En este capítulo hemos visto algunas formulaciones estadísticas más complejas y su aplicación particular sobre datos espaciales.

La clasificación es uno de los procesos de mayor importancia y, pese a estudiarse de forma habitual aplicado sobre imágenes, es de gran utilidad sobre cualquier tipo de datos. Tomando una serie de n capas ráster, la clasificación asocia cada celda a una clase dada, en función de los valores de dicha celda en esas capas.

Si en el proceso de clasificación se aporta algún tipo de información adicional sobre las características de las distintas clases, el proceso se conoce como clasificación *supervisada*. Si, por el contrario, se generan estas clases sin información adicional y simplemente buscando la mayor homogeneidad, el proceso se denomina clasificación *no supervisada*.

Otras formulaciones vistas son las relativas a la combinación de capas. A la hora de combinar varias de ellas, podemos realizar operaciones

aritméticas sencillas (mediante operaciones locales del álgebra de mapas) o aplicar otra serie de formulaciones más elaboradas. Metodologías como las *jerarquías analíticas* permiten establecer ponderaciones más correctas cuando el número de capas a combinar es elevado y resulta difícil asignarles pesos relativos. El significado de una capa en una operación de combinación puede ser distinto en función de si representa un factor más a considerar en la ecuación, o una restricción en el modelo.

Por último, hemos visto cómo el método de análisis de componentes principales permite reducir el número de variables con los que se trabaja en un modelo, reduciendo un conjunto de n capas a uno menor, tomando aquellas que explican la mayor variabilidad. Esto es de utilidad para establecer modelos de combinación entre capas, así como para reducir el volumen de datos en imágenes.

Parte IV

La tecnología

¿Cómo son las aplicaciones SIG?

22.1. Introducción

Las aplicaciones SIG son el elemento de trabajo básico dentro de todos aquellos que componen el concepto global de un SIG. Una aplicación SIG materializa todas las ideas vistas hasta el momento dentro de este libro, y es la herramienta fundamental para el trabajo con datos espaciales.

Dentro de la lógica evolución de toda tecnología informática, los SIG se han desarrollado de forma muy rápida y variada, adaptándose a una realidad, la de la propia información geográfica, también en constante evolución en todas sus vertientes. Por ello, la idea de aplicación SIG que podía encontrarse en un libro equivalente a este hace 10 o 20 años es bien distinta de la que hoy tenemos. De hecho, la concepción única de aquel entonces ya no es tal, y actualmente son muchas las formas en las que las aplicaciones SIG pueden presentarse.

Junto con la concepción «clásica» del SIG, todavía presente, existen una serie de otras tecnologías que han ido surgiendo paulatinamente, y que incorporan ideas y conceptos como los que ya conocemos de capítulos anteriores. En esta parte del libro se mostrarán todas ellas en detalle, definiendo así el panorama global de las aplicaciones SIG y los usos y funciones principales de cada una de dichas tecnologías.

Para comprender el papel que juegan las distintas formas de aplicaciones SIG que encontramos hoy en día, y que trataremos en los sucesivos capítulos, es necesario analizar primero la forma en que han ido conformándose dentro del entorno SIG.

22.2. La convergencia de las aplicaciones SIG

Una de las tendencias principales a lo largo de la evolución de los SIG es a la unión de otra serie de aplicaciones o elementos de estas, enriqueciéndose con conceptos y funcionalidades que, o bien encuentran en un SIG su aplicación a la información geográfica, o bien ya la tenían pero dentro de un marco aislado. El SIG actúa como elemento de unión de todas estas tecnologías, y engloba con carácter general a aquellas herramientas que de un modo u otro puedan emplearse para el análisis y tratamiento de datos espaciales.

Con esta filosofía, el concepto de SIG ha crecido desde sus orígenes mediante la incorporación de elementos propios de otras herramientas. Su crecimiento ha sido mayor que el de otros tipos de aplicaciones, ya que ha jugado un papel central y articulador, y en lugar de únicamente aportar conceptos a estas, en su mayoría ha tomado prestado de ellas. Dentro de las aplicaciones SIG actuales encontramos elementos que provienen, entre otros, de los siguiente ámbitos.

- Análisis de imágenes

- Diseño asistido por ordenador (CAD)
- Bases de datos
- Herramientas de diseño gráfico

Muchos de estos elementos ya se han comentado de uno u otro modo en secciones anteriores de este libro, ya que su importancia es más que notable.

Incluso dentro del propio ámbito SIG, las distintas aplicaciones han ido convergiendo paulatinamente. Las dos formas principales de almacenar la información geográfica, ráster y vectorial, conformaban originalmente también la base para las distintas aplicaciones, con escaso solape entre estas. Es decir, aquellas aplicaciones que podían manejar datos ráster y realizar operaciones con ellos, apenas tenían capacidades vectoriales o estas estaban por completo ausentes. Del mismo modo, las aplicaciones de corte vectorial no eran capaces de trabajar con datos ráster o, en todo caso, con algunas imágenes que podían representarse pero apenas analizarse.

Esta situación ha ido cambiando y, aunque en diferente forma, un SIG actual es capaz de trabajar con ambos tipos de datos con un nivel suficiente de funcionalidades. Poco a poco, todo el conjunto de tecnologías que han ido apareciendo dentro del entorno SIG se han ido extendiendo a las distintas aplicaciones, y aunque existen tipos bien definidos, estos no constituyen bloques estancos.

Así, por ejemplo, capacidades como el acceso a servicios remotos han evolucionado de forma similar a la gestión de datos ráster y vectoriales, en cuanto que han dejado de ser tecnologías exclusivas de una serie de aplicaciones para pasar a formar parte esencial del conjunto de estas. En el caso particular de estos servicios remotos, implicaron el desarrollo de servidores que eran mayoritariamente empleados desde aplicaciones Web. Con posterioridad, las aplicaciones de escritorio, más cercanas al concepto tradicional del SIG, han ido incorporando estas capacidades para ofrecer una funcionalidad similar a la de esas aplicaciones Web. En la actualidad, la integración de estos elementos va más allá, adaptando todas las restantes funcionalidades de esas aplicaciones de escritorio, muchas de las cuales no aparecen (todavía) en las aplicaciones Web, al trabajo con datos remotos.

De este modo, el trabajo actual con datos remotos se va integrando en los SIG como un elemento más, del mismo modo que ha sucedido con los distintos modelos de datos hasta alcanzar la situación actual, en la que se conciben como realidades distintas pero fundamentales y complementarias dentro de un SIG.

Veremos todo lo relativo al uso de datos remotos dentro del capítulo 24, también dentro de esta parte del libro.

22.3. La especialización de las aplicaciones SIG

Al mismo tiempo que las aplicaciones SIG iban incorporando funcionalidades e ideas de distintos ámbitos, surgían tecnologías y productos paralelos enfocados a un uso más concreto dentro de un determinado campo de aplicación. El crecimiento de los SIG que se produce como consecuencia de ese afán integrador da lugar a aplicaciones sólidas y completas, que resultan sumamente versátiles al tiempo que complejas. Siendo ya una tecnología base bien desarrollada, pueden comenzar a derivarse nuevas aplicaciones SIG que se asienten sobre esa base pero que no tengan tal carácter genérico, sino que concreten su campo de actuación y las tareas para las que están diseñados principalmente.

Por una parte, encontramos aplicaciones destinadas al uso en una determinada disciplina, en las que la aplicación conserva solo aquellas capacidades que resulten de mayor interés para el

objeto de esta. Las aplicaciones de este grupo pierden el carácter genérico y versátil del SIG, y normalmente integran tecnologías SIG dentro del marco de trabajo concreto de la disciplina correspondiente, aprovechando que en esta existe información geográfica susceptible de ser aprovechada mediante esas tecnologías SIG.

Por otra parte, encontramos modificaciones guiadas por los propios componentes de la herramienta, asignando más peso a elementos particulares del sistema SIG. De este modo surgen aplicaciones SIG dedicadas fundamentalmente a la gestión de datos, otras que se centran especialmente en el análisis, o bien aquellas en las que la visualización juega el papel fundamental. Sin olvidar que un SIG es ante todo un sistema, aparecen aplicaciones que concentran sus capacidades en un elemento de ese sistema. En lugar de entenderse la tecnología SIG como una aplicación que engloba a todo el sistema, se entiende ese sistema como un conjunto de aplicaciones más especializadas, cada una de las cuales constituye una de sus.

22.4. Tipos de aplicaciones

Con todo lo anterior, el panorama ante el que se encuentra hoy en día un usuario de SIG es sumamente complejo. Existen muchas aplicaciones distintas, y la dificultad de abordar su uso no es debida a su elevado número, sino a la gran cantidad de enfoques diferentes y conceptos distintos sobre los cuales estas se han desarrollado. En términos de tecnología, el mundo SIG es rico y variado, y resulta imposible tener un conocimiento completo de todos sus representantes.

Podemos distinguir tres grupos principales: herramientas de escritorio, repositorios de datos, y clientes y servidores que permiten en conjunto el trabajo remoto con todo tipo de datos SIG. Las herramientas de escritorio son la tecnología informática fundamental en el campo SIG. Los repositorios de datos y los clientes y servidores han ido cobrando día a día más importancia hasta convertirse en elementos fundamentales y muy representativos del mundo SIG actual. Ya conocemos bastante acerca de las bases de datos, puesto que debido a su relevancia las hemos desarrollado en capítulos anteriores del libro. Los clientes, por su parte, pueden presentarse de diversas formas, tanto como aplicaciones Web como integrados dentro de las herramientas de escritorio, aunque los estudiaremos junto a los servidores, agrupando así las tecnologías Web en un único bloque.

En los siguientes capítulos veremos las características de estos grupos, así como la relación existente entre ellos. Los repositorios de datos no tienen un capítulo propio dentro de esta parte, ya que hemos hablado de ellos en partes anteriores al tratar las bases de datos, pues así parecía más conveniente dada la importancia de estas y la necesidad de conocer algo más acerca de ellas antes de abordar otros temas como, por ejemplo, las consultas.

Juntos a estos tipos de *software*, encontramos otros de tipo SIG derivados de ellos, cuyo principal representante son las aplicaciones adaptadas a dispositivos móviles. Por la importancia que están cobrando en la actualidad estas últimas, detallaremos también sus características en un capítulo adicional.

Todas estos elementos conforman el panorama global de la tecnología SIG, con un conjunto de interrelaciones similar al definido esquemáticamente en la figura 22.1. Tanto clientes Web como herramientas de escritorio (en caso de que estas últimas tengan capacidades de cliente), acceden a los servidores para obtener datos y servicios. Los servidores, a su vez, toman datos de los repositorios de datos, al igual que pueden hacer las herramientas de escritorio para el trabajo con

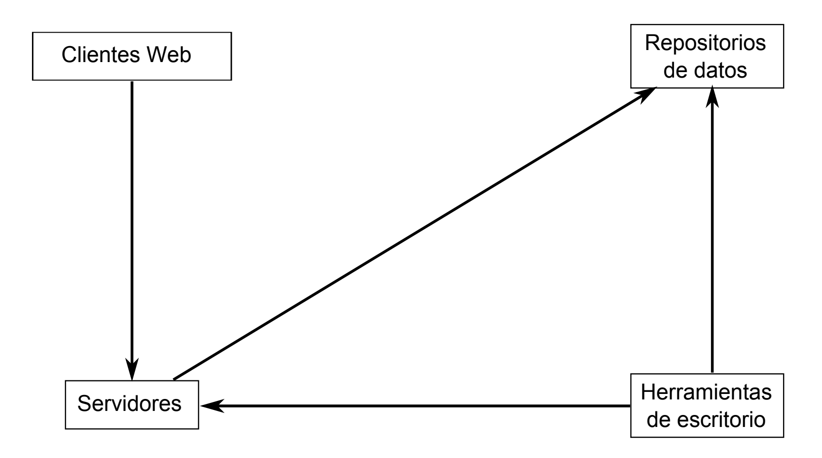


Figura 22.1: Clases principales de software SIG y relaciones entre ellas

datos locales, algo que los clientes Web no están pensados para hacer.

22.5. La adaptación de las aplicaciones SIG. El SIG como base genérica

Los SIG han crecido mucho desde su origen y, además de ampliar horizontes y mejorar el trabajo con ellos, han añadido numerosas funcionalidades adicionales. Como cabe esperar, un SIG actual no solo permite hacer las cosas mejor, sino que también permite hacer más cosas. Como herramienta rica en capacidades, un SIG puede entenderse como una aplicación preparada para responder a todas las posibles necesidades dentro del campo del análisis geográfico.

Sin embargo, la filosofía actual de las aplicaciones SIG es distinta a la existente en los primeros desarrollos, y el objetivo principal de un SIG hoy en día no es el de constituir una herramienta que contenga todas las funcionalidades que puedan necesitarse, sino una base sobre la que estas puedan construirse. Junto a las funciones básicas de edición, manejo de datos y análisis, un SIG permite la adaptación de estas a las necesidades concretas de cada trabajo, siendo así una herramienta versátil que puede tomar una u otra forma en función de las circunstancias particulares de cada uso.

La adaptabilidad de SIG es una de sus principales virtudes, y es la que permite que puedan desarrollarse útiles válidos para cada caso. Un SIG no es, por tanto, una herramienta cerrada con un conjunto de elementos suficiente para dar respuesta a todas las necesidades, y la obtención de una herramienta SIG final para un determinado trabajo no es un proceso único, sino un desarrollo en dos etapas.

La primera de estas etapas implica el desarrollo del propio SIG como tal, y la segunda concierne al desarrollo de elementos adicionales que completan la herramienta según las necesidades propuestas, apoyándose sobre los componentes fundamentales. Aunque muchos usuarios tendrán suficiente con un SIG en su forma original, muchos otros necesitarán desarrollos adicionales, o bien se beneficiarán de ellos al poder lograr mejoras sustanciales en comparación con el empleo del SIG básico.

Debido a este esquema de trabajo, el usuario SIG ha de ser en ocasiones un usuario técnico y cualificado, o bien ha de necesitar el concurso de alguien capaz de desarrollar sobre un SIG herramientas adicionales. La figura del programador SIG es importante dentro de un proyecto SIG, y hace que la gestión de la tecnología tenga la misma relevancia que la gestión de los datos o de cualquier otro de los restantes componentes globales de un SIG.

La idea de un SIG como herramienta base es especialmente patente en el caso de las aplicaciones de escritorio, las cuales concentran una gran mayoría del trabajo desarrollado dentro de un proyecto SIG. Esto las hace especialmente aptas a constituirse como herramientas básicas sobre las que se desarrollan modificaciones destinadas a responder a las necesidades del proyecto. No obstante, también otras aplicaciones SIG son susceptibles de jugar ese mismo papel.

En el caso de las aplicaciones Web, estas se adaptan para crear accesos particulares a unos datos concretos, de forma que pueden emplearse para dar acceso a la información geográfica a través de Internet, y hacerlo de una forma particular en cuanto a la apariencia y las funcionalidades ofrecidas. Los servidores se prestan de igual modo a ser adaptados en la medida de lo necesario.

Aunque la presencia de elementos para facilitar esa adaptabilidad (lenguajes de programación integrados, arquitecturas escalables, etc.) es general, la aparición de alternativas libres competitivas dentro del mercado del SIG ha potenciado más aún el desarrollo de herramientas adaptadas, al permitirlo en mayor grado.

22.6. Resumen

A partir de la concepción inicial de los SIG como aplicaciones bien definidas en las cuales se reunían las funcionalidades principales de estos, se ha desarrollado en la actualidad un amplio panorama de aplicaciones bien diferenciadas, las cuales podemos dividir en tres grupos principales: herramientas de escritorio, repositorios de datos, y clientes y servidores.

Estos tipos de aplicaciones se encuentran interrelacionados y se apoyan unos en otros para ofrecer todo el conjunto de capacidades actuales de los SIG.

Para llegar hasta este punto, los SIG han tomado elementos de otras aplicaciones, congregándolos en un único software. Al mismo tiempo, se han ido especializando en distintos ámbitos, dividiendo así el total de áreas de posible trabajo de este tipo de tecnologías.

En la actualidad los SIG forman una base genérica sobre la cual se construyen herramientas de análisis geográfico adaptadas a distintos fines.

Capítulo 23

Herramientas de escritorio

23.1. Introducción

El concepto clásico de un SIG es el de una aplicación completa en la cual se implementan herramientas para llevar a cabo las tareas básicas del trabajo con datos geográficos: creación o edición, manejo y análisis. Con esta filosofía fueron desarrollados los primeros programas SIG, especialmente para el tratamiento y análisis de datos geográficos y, posteriormente, para dotar a estos de mayor versatilidad. Se tienen así las herramientas de escritorio, que son aquellas que más se asemejan a la concepción original de los SIG.

A pesar de que existen, como vimos en el capítulo anterior, otras clases de aplicaciones SIG, los SIG de escritorio siguen manteniendo su posición como aplicaciones fundamentales, y hablar de un SIG implica por lo general hacerlo de una aplicación de escritorio antes que de otros tipos de aplicaciones. Por otra parte, las herramientas de escritorio son soluciones completas que cubren la totalidad de necesidades que se presentan en el desarrollo de proyectos SIG, y por ello constituyen las herramientas primordiales para llevar estos a cabo.

23.2. Funciones básicas

Podemos dividir las funciones básicas de un SIG de escritorio en cinco bloques: entrada y salida de datos, visualización, edición, análisis y generación de cartografía. Una aplicación de escritorio habitual presenta todas estas capacidades en cierta medida, aunque no necesariamente con el mismo nivel de implementación.

23.2.1. Entrada y salida de datos

Ya sabemos que los datos son una parte imprescindible de un SIG, y sin ellos no puede desarrollarse actividad con una aplicación SIG. Por esta razón, todas estas aplicaciones deben obligatoriamente implementar capacidades para leer datos y, opcionalmente, para guardarlos. Esta ultima es necesaria en el caso en que el SIG pueda generar nuevos datos geográficos (nuevas capas), pero no en aquellas aplicaciones sin capacidades de análisis o edición, donde su empleo no ha de crear nuevos datos.

Pese a ser de tal importancia, la implementación de las capacidades de entrada y salida es muy variable en unos u otros SIG. Una razón por la que esto sucede es el gran número de formatos de fichero distintos, la cual no favorece la interoperabilidad, como ya vimos en el capítulo 6. Así, cada SIG es capaz de abrir unos u otros formatos de archivo, y mientras que algunos tratan a todos ellos por igual, ciertas aplicaciones trabajan en un formato propio con carácter nativo, y son capaces de incorporar datos en otros formatos a través de extensiones o funciones de conversión entre estos y el formato particular del programa.

La existencia de librerías y componentes de acceso a datos en los que las aplicaciones SIG de escritorio pueden apoyarse mejora la conectividad entre estas y aporta cierta homogeneidad en este aspecto.

Otra diferencia importante en la gestión de los datos es la capacidad de conexión a bases de datos o servicios remotos como los que veremos en el capítulo 24[1]. Así, algunas aplicaciones tienen únicamente capacidad de abrir archivos, pero no de acceder a servicios remotos, mientras que otras pueden acceder a todo tipo de orígenes de datos. Las aplicaciones que acceden a servicios remotos entran dentro de la denominación de *clientes*. Los clientes Web que veremos en el capítulo 24 se conocen como clientes *ligeros*, mientras que las aplicaciones SIG de escritorio con estas capacidades se denominan clientes *pesados*, haciendo referencia a su mayor volumen y el mayor número de funcionalidades que contienen. La funcionalidad que ofrecen en relación al acceso a datos remotos es similar en ambos casos, y la trataremos en el citado capítulo 24 con detalle.

En general, las capacidades de acceso a servicios remotos están en función del enfoque de la aplicación, y aquellas fuertemente enfocadas al análisis no suelen presentar tales funcionalidades, mientras que aquellas más orientadas a la visualización o las herramientas globales más completas sí que las incorporan.

23.2.2. Visualización

La visualización es una función fundamental dentro de los SIG y del trabajo con cartografía en general. En este libro, por ejemplo, existe una parte completa dedicada a esta materia, pues su conocimiento resulta fundamental para poder aprovechar plenamente la información contenida en el dato geográfico.

Aunque existen SIG que no incorporan capacidades de visualización o estas no son muy avanzadas, la gran mayoría de las herramientas de escritorio incluyen un gran número de elementos para representar los datos geográficos con los que se trabaja. En ocasiones, interesa únicamente crear una representación de los datos, pero incluso cuando el trabajo con una herramienta SIG está enfocado a la realización de un análisis, la visualización y exploración visual de los datos de partida es un paso previo.

En general, la forma de operar con los elementos de visualización es muy similar entre soluciones SIG distintas y, a diferencia de lo que sucede con la implementación de otras funcionalidades, el manejo es prácticamente igual. Esto sucede no solo ya en las herramientas de escritorio que tratamos en este capítulo, sino también en las que veremos en el dedicado a las aplicaciones Web, las cuales incorporan también capacidades de visualización del mismo tipo. En ambos casos, las herramientas fundamentales presentan una estructura similar y unos conceptos base muy parecidos.

Esta estructura se compone fundamentalmente de un lienzo sobre el que se sitúan las distintas capas de información geográfica, y al que el usuario va añadiendo nuevas capas y editando la forma en la que estas se representan. Las capas se sitúan en un orden dado dentro del lienzo, lo que permite establecer una jerarquía de representación y así lograr el aspecto deseado.

Junto a este lienzo existen herramientas de navegación que permiten ampliar o reducir la escala, o bien modificar el encuadre (Figura 23.1). Asimismo, se presentan todas las funciones que permiten obtener las distintas formas de repre-

[1]Básicamente, estos servicios van a permitir acceder a datos geográficos que no están en nuestro ordenador, del mismo modo que accedemos a textos o imágenes a través de un navegador Web

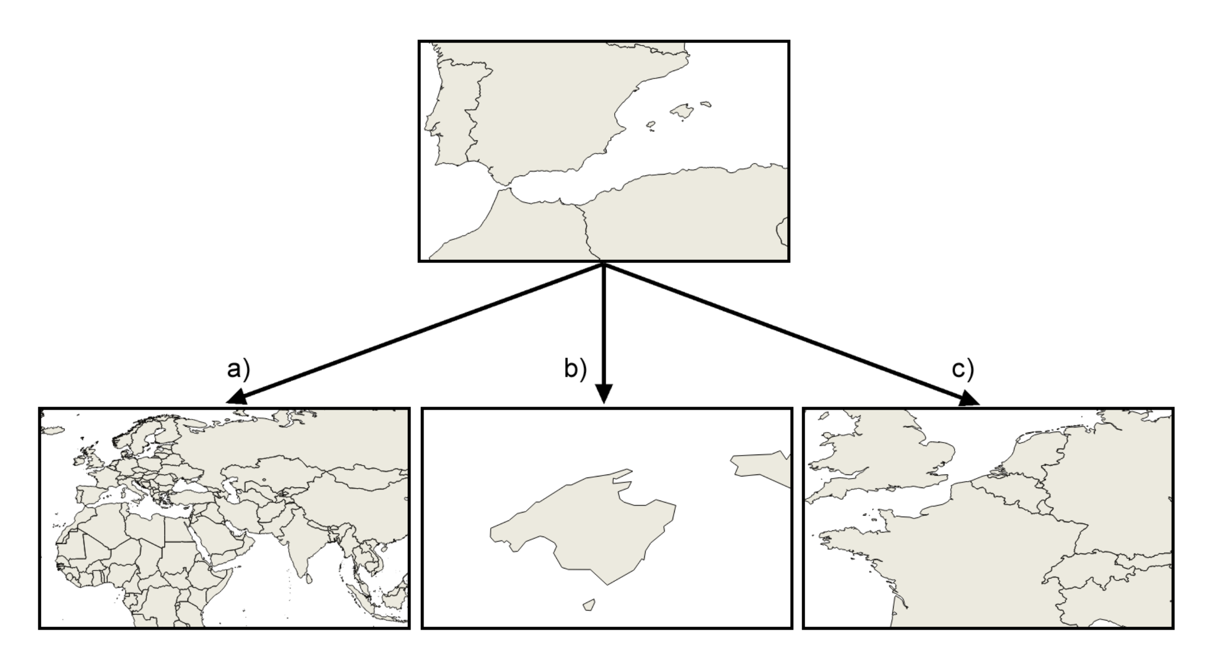


Figura 23.1: Herramientas de navegación fundamentales en el entorno gráfico de un SIG de escritorio. a) alejamiento (*zoom out*), b) acercamiento (*zoom in*), c) desplazamiento (*pan*)

sentación que veremos en la parte V, tales como el ajuste de colores o el establecimiento de etiquetas en función de los valores asociados a las distintas entidades, entre otros.

Estas capacidades convierte a los datos geográficos en un elemento activo, ya que, a diferencia de un mapa clásico donde no pueden modificarse sus características, en un SIG el usuario puede de forma rápida y sencilla elegir *qué* ve y *cómo* lo ve.

Como parece lógico pensar, la visualización ha evolucionado mucho desde los primeros SIG, y ha ido progresivamente adquiriendo nuevas capacidades, muchas de las cuales solo son posibles con los componentes gráficos de los ordenadores actuales. Así, además de ofrecer mayores posibilidades de personalización, el uso del SIG como herramienta de representación permite obtener resultados novedosos que añaden nuevas formas de explorar los datos geográficos y trabajar con ellos.

En el caso más habitual, la representación de una capa en un lienzo de un SIG es bidimensional, de la misma forma que se representa en un mapa impreso, lo cual se debe tanto a la mayor facilidad de implementación de este tipo de representaciones como a la mayor exigencia que otro tipo de representaciones presentan en lo referente al equipo (*hardware*).

No obstante, la presencia de visores tridimensionales está experimentando un gran crecimiento en los últimos años, y se van integrando paulatinamente dentro de los SIG para ofrecer una nueva forma de representación. Esta clase de capacidades gráficas, inalcanzables en términos de rendimiento para un equipo común hace unos años, son hoy día perfectamente utilizables en un ordenador de consumo habitual, y aportan una

nueva forma de trabajar con los datos geográficos.

Las funcionalidades de representación 3D aún no se integran completamente dentro del entorno de un SIG de escritorio completo, ya que se conciben por lo general como un elemento puramente visual y enfocado a la representación como tal, mientras que un lienzo 2D cumple su función tradicional de marco de trabajo sobre el que se desarrollan las restantes tareas del SIG como el análisis o la propia gestión de datos. Aun así, conforme las vistas tridimensionales se van convirtiendo en elementos habituales, los restantes componentes del SIG se van coordinando con ellas para darles a su vez la capacidad de servir como entornos de trabajo versátiles.

La presencia de una dimensión adicional hace que las herramientas de navegación sean más complejas en el caso tridimensional, existiendo ajustes relativos a la perspectiva, a los ángulos de visión o a la exageración del relieve, entre otros parámetros. Como puede verse en la figura 23.5, las herramientas de control de navegación de un visor 3D son sensiblemente más complejas que las herramientas base del caso 2D, presentadas en la figura 23.1.

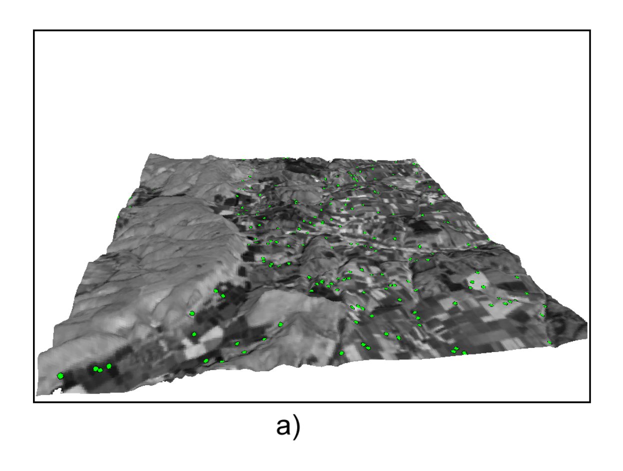
Esta mayor complejidad hace asimismo que puedan existir diversas formas en las que las capacidades de visualización 3D se presentan en un SIG. Las representaciones tridimensionales pueden ser simples representaciones en relieve de una capa (Figura 23.2a), o pueden incluir verdaderos elementos en tres dimensiones (Figura 23.2b). En el primer caso, la capa no contiene datos de elevación (puede ser por ejemplo una capa de usos de suelo), y la representación tridimensional se realiza basándose en la información de una capa adicional de elevaciones (que denominábamos de 2,5D por no poder representar todo tipo de formas en el espacio). La capa plana se representa en espacio «deformada» para ajustarse al relieve existente en la zona que representa.

En el caso de una representación tridimensional real, los objetos poseen información sobre su forma tridimensional, y junto con las coordenadas que delimitan su geometría plana (las geometrías básicas que conocemos: punto, línea y polígono), existen valores adicionales en el eje vertical. De este modo, pueden representarse entidades tales como edificios, o una ruta tridimensional que represente la trayectoria de un avión. Asimismo, y como puede verse en la figura, pueden añadirse elementos adicionales que aprovechan las capacidades de representación 3D, así como etiquetas o incluso elementos interactivos que también actúan en 3D (por ejemplo, para la selección de entidades).

Es importante recalcar que la visualización de una capa dentro de un SIG es independiente de la información que dicha capa contiene o la forma en que esta se almacena. El dato geográfico y su representación van por separado, y el dato en sí no define la representación, sino que únicamente sirve como apoyo para esta. Esto es particularmente cierto para el caso de capas vectoriales, así como para capas ráster que contengan un valor de tipo no gráfico, es decir, aquellas que no sean imágenes.

Una imagen tendrá el mismo aspecto en todos los SIG en los que se utilice, puesto que la información relativa a su representación se contiene, al menos en cierta medida, en la propia imagen. Los valores de la imagen representan una intensidad de luz y, si las bandas corresponden a la zona del visible, existe asimismo una información de colores. Otras imágenes pueden proceder del escaneo de un mapa o una fotografía impresa, y en ese caso recogen también los colores de cada píxel.

Aun en el caso de imágenes con más de tres bandas, en las que no existe una correspondencia


a)

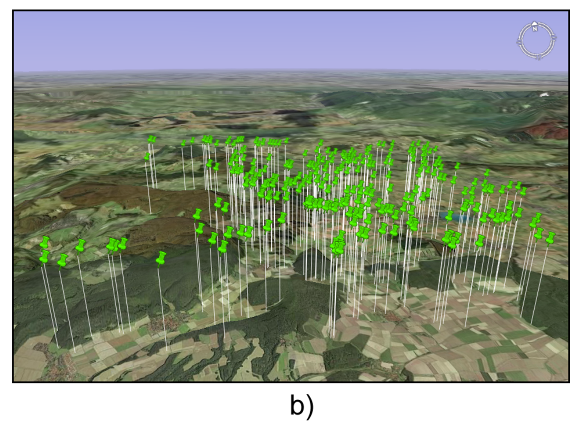

b)

Figura 23.2: Distintas formas de representación 3D de capas de datos geográficos

directa entre los valores y la representación (tal como vimos, por ejemplo, en la creación de representaciones en falso color en el apartado 17.4.1), puede existir una representación *por defecto* tomando, por ejemplo, las tres primeras bandas, y sigue siendo probable que estas imágenes también se vean de igual modo en uno u otro SIG (aunque luego ello no implica que dicha representación no pueda ajustarse a conveniencia en todos los casos).

En el caso de capas vectoriales o una capa ráster tal como un MDE, no existe ningún tipo de información acerca de la representación que acompañe al dato espacial en sí. Los datos necesitan de un esquema de asignación que los convierta en elementos visuales (colores, texturas, etc.), pero este esquema es ajeno al dato como tal.

La labor del SIG relativa a la visualización consiste en «interpretar» los datos y convertirlos en representaciones, y para ello se basa en esquemas definidos por el usuario. Estos esquemas pueden ser almacenados de forma que, en sucesivos usos de una capa de datos, esta se represente de una misma forma. No obstante, abrir la capa con una aplicación SIG distinta implicará en general perder las características definidas para su representación ya que, si bien los formatos de datos son relativamente interoperables, no así los formatos en que se almacenan los criterios de representación de esos datos.

Estándares para el almacenamiento de estilos como SLD(*Styled Layer Descriptor*), que veremos en detalle en el apartado 24.7.4, tienen como objeto solventar este problema.

23.2.3. Análisis

Si hubiéramos de ordenar cronológicamente las distintas funcionalidades que un SIG de escritorio presenta, probablemente el análisis fuera una de las primeras. Por encima de otras capacidades, los ordenadores han sido y son principalmente herramientas de cálculo capaces de realizar operaciones y computar resultados, y este ha sido uno de los usos fundamentales relativos al manejo de datos geográficos. Otros usos, tales como la visualización, pese a ser prácticamente imprescindibles hoy en día, estaban muy limitados en los primeros SIG. No ocurría así en el ámbito del análisis, ya que, aunque con capacidades como es lógico menores que las actuales,

los ordenadores ofrecían una potencia de calculo que los convertía en herramientas de análisis tan interesantes como en la actualidad.

La tendencia actual en los SIG es considerar las capacidades de análisis como herramientas modulares que se ejecutan sobre una plataforma base, la cual comprende las capacidades de visualización y entrada y salida de datos. Todas estas capacidades de análisis son independientes entre sí, aunque pueden coordinarse y emplearse en conjunto para alcanzar un resultado concreto. De otro modo, cada una de las formulaciones o algoritmos que vimos en capítulos anteriores aparece dentro del SIG como una herramienta individual que opera sobre una serie de capas y genera un resultado dado, tomando en muchas ocasiones esas capas de entre aquellas que se están representando en el SIG, e incorporando asimismo a dicha representación las nuevas capas generadas.

Las herramientas de análisis pueden aparecer igualmente como programas independientes, y el SIG de escritorio ser una herramienta aglutinadora que centraliza estas, facilitando su uso y la gestión de los datos implicados en los procesos de análisis.

Cuando las herramientas de análisis utilizan directamente la base del SIG en la que se encuentran a su vez las capacidades de visualización y manejo de datos, puede existir cierto grado de interactividad. Así sucede en las operaciones de consulta, las cuales vimos en el capítulo 12, ya que el usuario puede actuar sobre el lienzo para hacer una selección de modo gráfico. Más aún, esa selección puede condicionar los posteriores análisis sobre la capa cuyas entidades se seleccionan, ya que los procesos que operen sobre ella pueden restringir su alcance a aquellas entidades seleccionadas.

En caso de no existir este tipo de interacción entre elementos de análisis y elementos de visualización y exploración de datos, los procesos de análisis suelen constituir utilidades autocontenidas que simplemente toman una serie de datos de entrada, realizan un proceso en el que el usuario no interviene, y finalmente generan un resultado con carácter definitivo. Este resultado podrá ser posteriormente visualizado o utilizado como entrada para un nuevo análisis.

A modo de ejemplo, podemos analizar el caso particular del cálculo de una ruta que conecte una serie de puntos a través de una red (los fundamentos de este análisis los vimos en el apartado20.7).

Una de las formas de implementar este análisis es aquella que requiere del usuario la introducción de la información necesaria (una capa de líneas con la red viaria y otra de puntos con los puntos de inicio, paso y llegada correspondientemente ordenados) como parámetros que el proceso toma y en función a los cuales se genera una nueva capa. El proceso es una tarea perfectamente definida, con unas entradas y una salidas, y tras la selección de unas capas de entrada se genera un nuevo resultado en forma de otra capa. Este proceso puede implementarse de forma aislada del SIG, aunque coordinada con él para lograr mejores resultados y una utilización más sencilla.

Otra forma con un enfoque distinto sería presentando un proceso interactivo en el cual se introduce como único parámetro inicial la red viaria. Posteriormente, el usuario puede operar sobre el lienzo en el que esta se encuentra representada para ir definiendo y editando la lista de puntos de paso. El cálculo se va efectuando cuando se produce algún cambio en la misma y debe aplicarse de nuevo el algoritmo pertinente para adaptar el resultado —la ruta óptima— a ese cambio.

Este tipo de formulaciones interactivas son más intuitivas y agradables de usar, pero en realidad menos productivas de cara a un trabajo dentro de un proyecto SIG. Aparecen por ello en

aquellas funciones que tienen una mayor componente visual o, especialmente, en las que representan un análisis puntual que se realiza de forma común como algo individual. Estos son los análisis que se implementan en las aplicaciones que veremos más adelante como parte del grupo de SIG enfocados a la exploración visual de datos geográficos, que además de esta proveen alguna serie de utilidades de análisis, pocas en general, sin que estas estén concebidas para un trabajo completo en un proyecto SIG de cualquier índole.

En un proyecto SIG de cierto tamaño, lo más común es que la fase de análisis, a la que seguirá una fase también compleja de preparación e interpretación de resultados, y previa a la cual se ha debido llevar a cabo la preparación de los datos de partida, comprenda no uno, sino muchos análisis distintos. Estos análisis, a su vez, no son independientes, sino que están relacionados entre sí, y lo más habitual es que definan un flujo de trabajo que comienza en los datos de partida y desemboca en los resultados finales a través de una serie de procesos.

Por su naturaleza, tanto los datos espaciales como los procesos en los que estos intervienen se prestan a formar parte de estos flujos de trabajo más o menos complejos, y es por ello que en los SIG actuales una funcionalidad básica es la creación de tareas complejas que permiten simplificar todo un proceso de muchas etapas en una única que las engloba a todas. La forma anteriormente comentada en que aparecen las formulaciones dentro de un SIG, de forma atomizada y modular, facilita la creación de estos «modelos» a partir de procesos simples.

Para entender esta idea, podemos ver un ejemplo aplicado. La extracción de una red de drenaje en formato vectorial a partir de un MDE requiere una serie de procesos, a saber (para repasar los fundamentos de cada uno de estos procesos y comprender mejor la operación global, puede consultarse el capítulo 16, donde se describieron en su momento, así como la sección 18.2.1):

1. Eliminación de depresiones del MDE.
2. Cálculo de una capa de área acumulada a partir del MDE corregido
3. Extracción de capa ráster con la red de drenaje a partir de la capa anterior y un valor umbral
4. Vectorización de la capa resultante del paso anterior

Lo anterior podría simplificarse si se agruparan en una sola operación todos los procesos anteriores, de forma que tomando dos datos de entrada (un MDE y un umbral), se realizara todo el proceso de forma continua. Las capacidades de creación de modelos que implementan los SIG de escritorio (en particular, los más enfocados al análisis) permiten crear nuevos procesos compuestos utilizando entornos gráficos intuitivos, y estos procesos pasan a formar parte del conjunto de ellos de que dispone el SIG, permitiendo que cada usuario «resuma» en procesos unitarios una serie de operaciones que, de otro modo, deberían realizarse de forma individual con el coste operacional que ello implica.

Una vez que se ha creado un proceso compuesto, este puede aplicarse sobre nuevos datos de entrada, reduciéndose así el tiempo y la complejidad con que nuevos parámetros de entrada pueden ser procesados según el esquema de trabajo definido en dicho proceso.

Debe pensarse que el proceso presentado como ejemplo es muy sencillo y únicamente implica cuatro operaciones encadenadas de forma lineal. Un proceso de análisis habitual puede contener muchas más operaciones individuales, y estas disponerse de forma más compleja, con dependencias de distinta índole entre sus resultados.

La imagen 23.3 muestra el aspecto de uno de tales procesos implementado en un software GIS con capacidad de de creación de modelos. El

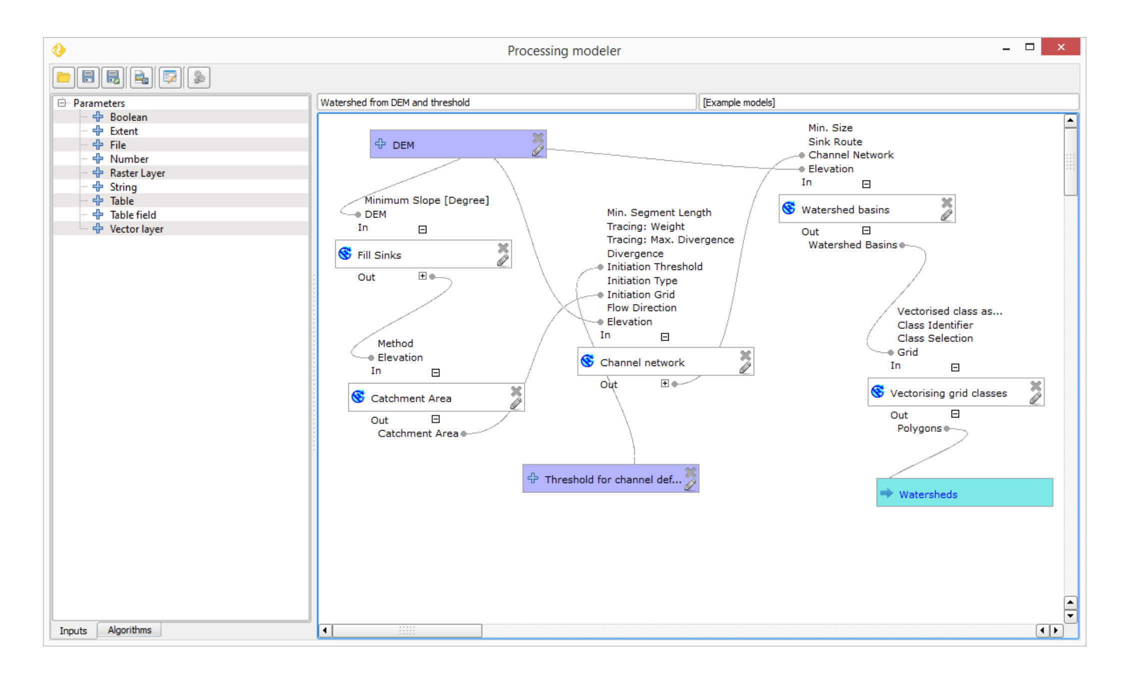


Figura 23.3: Esquema de un proceso complejo creado a partir de operaciones simples de análisis con datos SIG.

proceso representado en la figura 7.6 también es un ejemplo de otro tipo de análisis que puede adaptarse ventajosamente en este tipo de herramientas de modelización.

Asimismo, las herramientas SIG que contienen funcionalidad de análisis suelen permitir el acceso a estas a traves de lenguajes de *scripting*, lo cual facilita la creación de flujos de trabajo y la automatización de rutinas complejas de análisis. Si bien este trabajo no se realiza de un modo gráfico e intuitivo como en el ejemplo mostrado anteriormente, y requiere mayores conocimientos, la flexibilidad y potencia que ofrece es mucho mayor.

23.2.4. Edición

Los datos geográficos con los que trabajamos en un SIG no son una realidad estática. La información contenida en una capa es susceptible de ser modificada o corregida, y las funciones que permiten estas tareas son importantes para dotar al SIG de versatilidad. Sin ellas, los datos espaciales pierden gran parte de su utilidad dentro de un SIG, ya que se limitan las posibilidades de trabajo sobre estos. Las funcionalidades de edición son, por tanto, básicas en una herramienta de escritorio.

Las operaciones de edición pueden emplearse, por ejemplo, para la actualización de cartografía. Como vimos en el capítulo 6, una de las ventajas de los datos digitales frente a los analógicos es la mayor facilidad de actualización. Así, si una entidad en una capa vectorial modifica su geometría, no es necesario rehacer todo un mapa, sino simplemente editar ese elemento. A lo largo del desarrollo de un proyecto SIG, es muy probable que sea necesario editar de un modo u otro algún dato espacial, bien sea para corregirlo, ampliarlo, mejorarlo o sencillamente adaptarlo a las necesidades del propio proyecto.

Además de la modificación de una capa ya existente, las herramientas de edición de un SIG de escritorio se emplean igualmente para la creación de capas nuevas, que pueden crearse a partir de la digitalización de imágenes como vimos en el capítulo 6, o bien en base a cualquier otra capa de la que dispongamos.

Aunque las tareas de edición más habituales son las relacionadas con la edición de geometrías, no es esta la única edición que puede realizarse dentro de un SIG. Podemos distinguir las siguientes formas de edición:

- Edición de geometrías de una capa vectorial
- Edición de atributos de una capa vectorial
- Edición de valores de una capa ráster

Las herramientas destinadas a la edición de entidades geométricas heredan sus características de los programas de diseño asistido por ordenador (CAD), cuya funcionalidad principal es precisamente la edición de elementos gráficos. Estas incluyen la adición o eliminación de nuevas geometrías, la modificación de ellas editando sus puntos (recordemos que toda entidad vectorial se reduce a un conjunto de puntos en última instancia), así como otras operaciones geométricas básicas. En la sección 6.5.1 vimos algunas de ellas a la hora de tratar la calidad de la digitalización en pantalla.

Otras funciones de edición que encontramos son las que permiten simplificar algunas tareas, tales como la división de un polígono. Otras funcionalidades similares incluyen la eliminación automática de polígonos espúreos (véase 19.3.5), o el ajuste automático entre entidades.

En general, el número de funcionalidades es sensiblemente menor que en el caso de los programas CAD, ya que gran parte de ellas no tiene aplicación directa en el caso de trabajar con datos geográficos. No obstante, también aparecen herramientas adicionales, como sucede en el caso de que se registre información topológica, lo cual ha de considerarse por igual en el proceso de edición.

Junto con las propias geometrías, toda capa vectorial tiene asociado un conjunto de atributos, y estos deben poder editarse también desde el SIG. De hecho, la adición de una nueva geometría a una capa vectorial no está completa hasta que no se añaden igualmente sus atributos.

La edición de toda la información alfanumérica relacionada con las distintas entidades se realiza en un SIG a través de elementos tomados del ámbito de las bases de datos, siendo esto en general válido tanto en lo referente a las interfaces como en el propio acceso a datos. Las operaciones de edición de atributos abarcan tanto la modificación de valores sencillos como la de la propia estructura del conjunto de atributos (adición o eliminación de columnas —campos— en la tabla correspondiente). Por supuesto, la edición de atributos y de geometrías esta íntimamente relacionada.

Por último, la edición de capas ráster es mucho menos frecuente, y una gran mayoría de SIG no permiten la modificación directa de los valores de las celdas. Las operaciones del álgebra de mapas permiten modificar los valores de una capa y obtener nuevas capas con esos valores modificados, pero editar directamente un valor de celda del mismo modo que se editan el valor de un atributo o la posición de un punto de una capa vectorial no es una funcionalidad tan habitual.

23.2.5. Generación de cartografía

A pesar de que la representación de las distintas capas de datos espaciales en un lienzo es suficientemente potente a efectos de explorar visualmente la información que estas contienen, la mayoría de los SIG incorporan capacidades de creación de cartografía impresa, generando un documento cartográfico que posteriormente pue-

de imprimirse y emplearse como un mapa clásico. Las razones para la existencia de tales funcionalidades son muchas, pero la principal sigue siendo la necesidad general que aún existe de apoyarse en esa clase de documentos cartográficos.

Aunque la representación dentro de un SIG ofrece posibilidades mayores (cambio de escala, ajuste de los parámetros de visualización, etc.), disponer de una copia «estática» de cada bloque de información con el que se trabaja en un SIG es una necesidad ineludible. Más que una capacidad necesaria para la presentación adecuada de la información cartográfica, la generación de cartografía impresa es en muchos casos la principal razón para el uso de un SIG.

Fundamentalmente, estas capacidades permiten la composición de documentos cartográficos de acuerdo con un diseño dado, y la impresión directa de estas en algún periférico tal como una impresora común o un *plotter* de gran formato. En la elaboración de dicho diseño, pueden emplearse todos los elementos que habitualmente podemos encontrar en un mapa: el propio mapa en sí (la representación de la información geográfica), leyenda, título, escala, etc. Con estos elementos, se crea una versión autocontenida de la información geográfica, que puede ya emplearse de modo independiente del SIG.

Las funciones de diseño que se implementan por regla general en un SIG son similares a las que pueden encontrarse en un software de maquetación genérico, permitiendo la composición gráfica del documento general y el ajuste de los distintos elementos que lo forman. Se diferencian de aquellas en el hecho de que toman en consideración la naturaleza cartográfica de los elementos que componen el documento, y por ello disponen de funcionalidades para manejar correctamente parámetros tales como la escala de representación cuando se rediensionan dichos elementos.

Una de las funciones más interesantes de generación cartográfica en un SIG es la automatización del proceso y la simplificación de la producción de grandes volúmenes de cartografía. Por una parte, todas las herramientas de escritorio capaces de producir cartografía son a su vez capaces de «reutilizar» diseños, de tal modo que si un conjunto de mapas tienen unas características comunes (por ejemplo, una misma disposición de sus elementos), no es necesario elaborar todos ellos desde cero.

Esto permite, por ejemplo, crear una serie de mapas de una misma zona conteniendo cada uno de ellos información sobre distintas variables. A partir de un conjunto de capas, se elabora el diseño de un mapa y este se alimenta de dichas capas, creando mapas independientes que reflejan estas por separado o en distintas combinaciones. Esto simplifica notablemente el proceso, ya que el diseño ha de realizarse una única vez, al tiempo que se garantiza la uniformidad de los distintos resultados.

Otra aplicación en esta línea es la generación de una serie de mapas que cubren en su conjunto una amplia extensión, fragmentando esta en unidades. La gestión de los encuadres para cada una de esas unidades, o la creación de un mapa guía en cada caso que localice la hoja concreta dentro de la extensión global del conjunto, ambas pueden automatizarse junto con las restantes operaciones de diseño. De este modo, la producción de toda una serie cartográfica se simplifica en gran medida, siendo el SIG una herramienta que supone un gran avance en términos de productividad en este tipo de tareas.

La figura 23.4 muestra un ejemplo de lo anterior.

Estas posibilidades surgen de la separación existente en un SIG entre los datos espaciales y el diseño del documento cartográfico que los contiene, del mismo modo que ya vimos existe entre

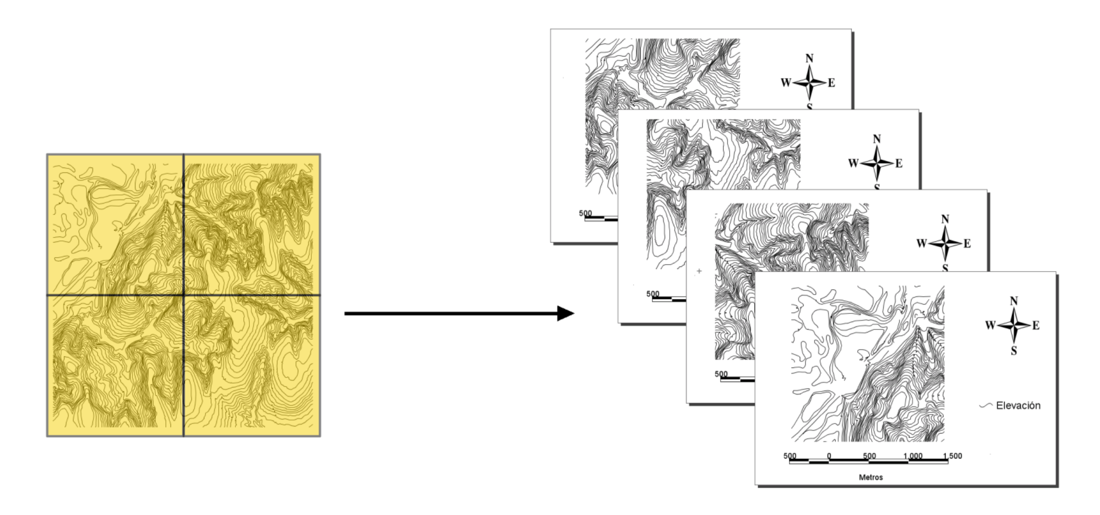


Figura 23.4: La automatización de las tareas de creación cartográfica permite simplificar la producción de grandes volúmenes de cartografía, como por ejemplo al dividir un área geográfica en una serie dada de mapas.

datos y parámetros de representación a la hora de visualizar los primeros.

23.3. Tipos de herramientas de escritorio

No todas las aplicaciones de escritorio presentan las funcionalidades anteriores de igual modo. Es frecuente que ciertos Sistemas de Información Geográfica tengan una fuerte componente de análisis, pero que otras de las funciones principales, como por ejemplo la edición, no se presenten tan desarrolladas.

Un caso particular es el de aquellos SIG de escritorio que centran la gran mayoría de sus capacidades en el terreno de la visualización, permitiendo un uso de los datos geográficos similar al que corresponde a un mapa clásico, donde el trabajo con este se basa fundamentalmente en el análisis visual.

Comenzaremos por estos últimos para dar un breve repaso a los principales tipos de aplicaciones de escritorio en función de sus capacidades.

23.3.1. Visores y exploradores

Las aplicaciones SIG de escritorio cuya función principal es la visualización se conocen generalmente como *visores* o *exploradores*, y en la actualidad representan una fracción importante del conjunto total de herramientas SIG de escritorio.

En ocasiones, se trata de aplicaciones en el sentido habitual, las cuales presentan capacidades reducidas de análisis y edición, y cuyo objetivo no es otro que el permitir la visualización de cartografía, sin incorporar las restantes posibilidades del SIG. En otros casos, son versiones simplificadas de soluciones SIG de escritorio más complejas, desarrolladas como alternativas más asequibles (en términos de dificultad de manejo y aprendizaje, y también a veces en términos de coste).

Una forma también habitual en la que se presentan los exploradores son como herramientas de apoyo a unos datos espaciales particulares. Para entender este tipo de enfoque, debe pensarse que un mapa clásico puede visualizarse de igual

modo con independencia del uso que se le pretenda dar y de la experiencia y formación de quien lo usa. Con un conjunto de datos espaciales en forma de una o varias capas, no sucede lo mismo, ya que estos datos no son un elemento «visual» de por sí. Es necesario utilizar un SIG para poder visualizarlos.

Un usuario experimentado no encontrará problemas en manejar un SIG de escritorio complejo, pero un usuario con poca experiencia que lo único que desee sea «ver» la cartografía y explorarla visualmente (del mismo modo que un excursionista casual puede querer emplear un mapa topográfico) encontrará el entorno de ese SIG demasiado complejo y con elementos que, en su mayoría, no le son necesarios. Con la disponibilidad creciente de cartografía y la popularización de las tecnologías SIG, este tipo de usuarios ha crecido notablemente, y las aplicaciones adaptadas a sus necesidades han ido apareciendo y popularizándose de forma progresiva.

Así, existen visores que ocupan un papel secundario como parte de un producto compuesto que incluye al propio programa y a los datos en sí. Ejemplos muy claros y muy populares son aplicaciones como Google Earth (figura 23.5) , que permite que cada usuario incorpore su propia información para visualizar esta, pero cuyo mayor interés es el acceso a una enorme base de datos de imágenes de satélite con cobertura global. De esta forma, la aplicación puede utilizarse para explorar una área deseada, sin necesidad de disponer junto con ella de datos para dicha zona, puesto que por defecto la aplicación accede a una base de datos de imágenes que van indisolublemente unidos a ella.

El usuario puede también añadir sus propias capas y usar estos visualizadores de la misma manera que emplea las capacidades de visualización de un SIG de escritorio más rico en funcionalidades, pero parte del interés de la aplicación no está solo en sus capacidades, sino en los datos que tiene vinculados y que permiten emplear estas.

En otros casos, un organismo o empresa puede generar un conjunto de capas bien a partir de algún tipo de análisis o de cualquier otra metodología, y opta por distribuir estas acompañadas de un visor que permita un primer acceso a los datos. Todas esas capas podrán ser empleadas dentro de un SIG que soporte los formatos de archivo en el que estas se hayan almacenado, pero aquellos usuarios que no dispongan de un SIG podrán igualmente efectuar consultas básicas y explorar la cartografía haciendo uso del visor incorporado.

En líneas generales podemos enumerar las siguientes características de los visores:

- Interfaz simple en la que tienen un peso mayoritario las herramientas de navegación.
- Capacidades de lectura de datos, pero no de escritura.
- Reducidas o nulas capacidades de edición y análisis.
- Enfocadas a usuarios no especializados.

Encontramos dos grupos básicos de visores, en función de qué tipo de visualización principal incorporan: planos y tridimensionales (globos). Mientras que una aplicación SIG de escritorio completa puede presentar los dos tipos de representación, y sobre ambas implementar las restantes funcionalidades tales como la edición o el análisis (aunque, como ya dijimos, en mayor proporción sobre las vistas bidimensionales), los visores habitualmente reducen sus capacidades de representación a una de estas variantes.

Los visores bidimensionales, aunque sin alcanzar el enfoque especializado de una aplicación completa, se orientan más al usuario con cierto conocimiento del ámbito SIG, mientras que la tendencia en los tridimensionales es a ofrecer herramientas de acceso a datos geográficos con una apariencia atractiva. Ello, no obstante, no implica que estos visores carezcan de utilidad en el ám-

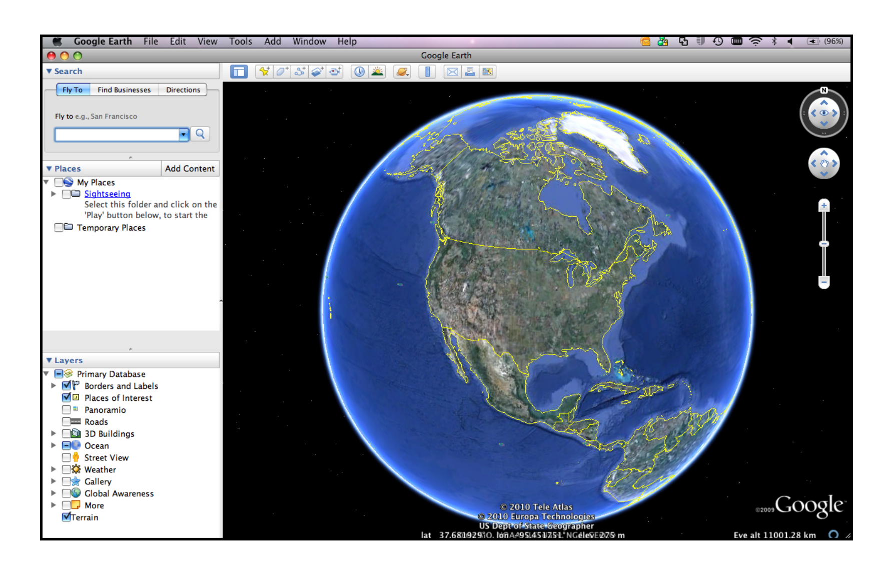


Figura 23.5: Aspecto de un globo o visor tridimensional (GoogleEarth).

bito científico, siendo igualmente herramientas válidas para todo tipo de investigación o trabajo que incorpore cierta componente geográfica. De hecho, la popularización de este tipo de visores ha supuesto un gran acercamiento de los datos geográficos y las capacidades SIG a toda una amplia comunidad de usuarios, incluyendo los del mundo científico, permitiéndoles realizar sus trabajos de forma más adecuada. Esto es especialmente cierto con aquellos visores que se hallan vinculados a bases de datos particulares, como ya se ha comentado, ya que permiten explotar los datos de dichas bases y proporcionar a todo usuario un sustrato de información geográfica sobre la que trabajar.

23.3.2. Soluciones de escritorio completas

La aplicación SIG más habitual, y la que constituye la herramienta básica para el desarrollo de un proyecto SIG, es aquella que reúne en un único producto todas las funciones básicas que hemos visto en este capítulo. Con las lógicas diferencias en cuanto al grado de funcionalidad de estas según su enfoque, una aplicación SIG de escritorio completa debe permitir la lectura de los datos, la creación y modificación de estos con sus capacidades de edición, su visualización, la realización de análisis con ellos y la generación de resultados cartográficos ya sea a partir de los datos originales o de datos generados en los procesos de análisis. Con todas estas capacidades, una herramienta SIG de escritorio constituye una solución completa para todo tipo de proyectos SIG, y puede dar respuesta a todas las necesidades que en ellos se presentan.

Como ya vimos en el capítulo introductorio de esta parte, y también en el capítulo 8 dedicado a las bases de datos, la forma de abordar la implementación de las distintas capacidades ha ido variando a medida que se iban desarrollando los SIG. En la actualidad, encontramos tanto SIG de escritorio que implementan en una sola aplicación central todas las funcionalidades base, como grupos de aplicaciones muy interrelacionadas que implementan por separado cada una de dichas funcionalidades. Tanto en uno como

en otro caso, existen elementos sobre los que las distintas herramientas del SIG se apoyan, especialmente en lo relativo al acceso a datos.

Pese a que incorporan toda la gama de funcionalidades base, las soluciones de escritorio completas no cubren las necesidades de todo usuario de SIG. La convergencia a la que tienden este tipo de aplicaciones, comentada en el capítulo introductorio de esta parte del libro, y como ya entonces se mencionó, no se ha logrado completamente. El problema no es ya un problema de integración de tecnologías, sino una dificultad relativa a la gran amplitud de la ciencia de los SIG. Resulta imposible reunir en una sola herramienta todas las capacidades que un SIG de escritorio puede incluir, y es por ello que todas las soluciones de escritorio presentan algún tipo de especialización, dando prioridad a algún área respecto a las restantes.

Una división que perdura, aunque en mucho menor medida que en los primeros SIG, es la existente entre SIG ráster y SIG vectoriales, especialmente en lo relativo al análisis. Un usuario que trabaje con datos espaciales como los correspondientes, por ejemplo, a información catastral, utilizará para su labor un SIG de escritorio distinto al que usara alguien cuyo trabajo implique mayoritariamente la realización de análisis del terreno o la creación de modelos geográficos, a pesar de que ambas soluciones probablemente incorporen capacidades tanto en el ámbito ráster como en el vectorial. El alcance de estas, no obstante, será distinto en cada caso.

Incluso en caso de presentar una orientación principalmente hacia los datos de tipo ráster, existen también enfoques distintos dependiendo principalmente del tipo información que se vaya a manejar. Una división fundamental es la existente entre aquellas aplicaciones destinadas al manejo de imágenes y aquellas cuyo elemento principal de trabajo son las capas ráster con otro tipo de valores, tales como MDE o capas similares.

Las imágenes, especialmente si se trata de imágenes de satélite, van a constar de una serie de bandas cuyo número puede ser muy elevado, lo que requiere unas herramientas particulares para su manejo. Asimismo, gran parte de las funciones que vimos en el capítulo 17 tales como las relativas a la corrección de imágenes, no tienen aplicación para otro tipo de capas (no tiene sentido aplicar una corrección geométrica a un MDE o una capa de pendientes, ya que estos datos necesariamente provienen de fuentes ya conveniente corregidas), por lo que en cierto modo pueden considerarse especificas de este campo, el de las imágenes, si bien es cierto que se trata de un campo de gran amplitud.

Por el contrario, capas como el propio MDE u otras similares solo contienen una única banda, y el tipo de operaciones que se desarrollan sobre ellos es bien distinto. Con objeto de simplificar estas operaciones, la estructura de estas aplicaciones ha de enfocarse hacia alguna de estas variantes, dándole prioridad sobre la otra. Por ello, las herramientas de escritorio que se orientan al trabajo con imágenes incorporan en general pocas o nulas herramientas en áreas como el análisis del terreno, mientras que aquellas que sí tratan estos análisis no incluyen salvo las funciones más simples para el manejo de imágenes (realces, ajuste de contraste, etc.), pero no las más específicas.

En realidad, una aplicación de escritorio global que cubriera todas estas funcionalidades no sería práctica desde el punto de vista de su uso, pues sería excesivamente compleja. Es poco probable, igualmente, que un mismo usuario requiera un entorno profesional productivo en todas ellas, y es más habitual que centre su trabajo en un área concreta.

23.4. Resumen

Las herramientas de escritorio son la forma más clásica de los SIG. Entendemos como tales a aquellas herramientas ciertamente complejas que permiten llevar a cabo las tareas básicas de un SIG en sentido tradicional, como son el manejo de datos espaciales y el trabajo con los mismos.

Podemos distinguir cuatro funcionalidades básicas que aparecen representadas en mayor o menor medida en un SIG de escritorio: visualización, edición, análisis y generación de cartografía.

En función del grado de desarrollo e implementación en que las anteriores funcionalidades se encuentren en un SIG de escritorio, distinguimos distintas formas de estas herramientas. La división más genérica es aquella que distingue las herramientas pensadas para un trabajo completo en todas las distintas fases de un proyecto SIG de aquellas orientadas a la representación y exploración visual de los datos geográficos. Estas últimas representan un enfoque más reciente, y en la actualidad están contribuyendo de manera muy notable a la expansión de las tecnologías SIG fuera del ámbito más especializado.

Capítulo 24

Servidores remotos y clientes. Web Mapping

Olaya, Víctor; Turton, Ian; Fonts, Oscar

24.1. Introducción

Del mismo modo que podemos acceder a otros tipos de información a través de Internet o de una red local, también podemos emplear esta para acceder a información geográfica y trabajar con ella dentro de un SIG. En el contexto actual, no puede dependerse en un SIG únicamente de datos locales en forma de archivos en el mismo ordenador en el que se trabaja, sino que es necesario poder operar con datos remotos. Las redes son la vía para la difusión de todo tipo de información, entre ella la información geográfica.

Los datos espaciales pueden ofrecerse a través de una red de la misma manera que se ofrecen otro tipo de datos como imágenes o texto en una pagina Web. Para que en este proceso se maximicen las posibilidades que esos datos ofrecen, es necesario disponer de tecnologías adaptadas, particularizadas al tipo concreto de datos que se maneja y los posibles usos que pueden darse.

Estas tecnologías son variadas y, como cabe esperar, han evolucionado paralelamente a otras basadas en la Web, añadiendo progresivamente elementos tales como una mayor interactividad o flexibilidad. Las páginas estáticas que formaban Internet hace unos años, muy limitadas en cuanto a sus posibilidades, han dado paso a lo que hoy se conoce como Web 2.0, donde encontramos *blogs*, *wikis* y otros tipos de páginas con capacidades mucho mayores y que permiten al usuario un trabajo muy distinto.

Una evolución similar han seguido las aplicaciones de la Web relacionadas con la información geográfica, que han ganado en riqueza hasta el estado actual donde pueden llegar a ofrecer casi tantas funcionalidades como un SIG de escritorio. Los mapas estáticos que constituían los primeros elementos con componente geográfica en la Web han evolucionado hasta verdaderas aplicaciones que pueden convertir un navegador en una plataforma SIG completa. En su avance, las tecnologías Web van tomando elementos que ya conocemos de los SIG de escritorio, con objeto de trasladar toda su potencia al entorno de Internet, y uniéndola así con las capacidades que la red tiene como espacio común de actividad y conocimiento.

Aunque el objetivo final sea trasladar los SIG de escritorio a la red, las tecnologías necesarias distan bastante de las tecnologías SIG en sentido clásico, de la misma forma que, aun trabajando con un tipo de datos similar, un procesador de textos se diferencia mucho de un navegador Web.

Fundamentalmente, estas tecnologías Web han de responder a dos necesidades principales: servir un elemento a través de la red y tomar este

para emplearlo. Es decir, tomar y recibir el elemento que es objeto de interés. Distinguimos así los conceptos de *servidor* y *cliente*, que debemos ver con algo más detalle antes de continuar.

24.2. ¿Cómo funciona Internet?

Estamos acostumbrados a utilizar Internet a través de aplicaciones tales como navegadores Web, y en muchos casos desconocemos cómo se realiza ese proceso tan cotidiano hoy en día. Los fundamentos que residen detrás de la consulta de una simple página Web son esencialmente los mismos que vamos a encontrar para el caso de las tecnologías SIG en la red, por lo que es necesario conocerlos al menos someramente para poder entender el proceso que tiene lugar cuando empleamos una tecnología SIG en Internet.

Cuando consultamos una página Web existen tres elementos fundamentales que entran en juego: la propia red que hace de nexo entre sus elementos, nuestro ordenador que es el que realiza la petición de consulta, y la máquina donde se encuentra almacenada esa página que queremos consultar.

Conocemos como *servidor* al elemento encargado de *servir* algún tipo de contenido. En el ámbito SIG, se trata fundamentalmente (aunque no con carácter exclusivo) de datos geográficos, que constituyen el principal producto que se distribuye a través de la red dentro de nuestro campo. En el ejemplo anterior, la máquina que contiene la página de interés es el servidor. También se conoce como servidor el programa que, residiendo en esa máquina, interpreta la petición y la procesa, sirviendo así la página.

El *cliente* es responsable de *pedir* ese dato al servidor, tomarlo y trabajar con él. Nuestro navegador Web es el cliente en este caso, ya que es el que realiza la petición. Para ello, basta con introducir la dirección Web[1] correspondiente en la barra de direcciones del navegador. Al hacer esto, proporcionamos una serie de datos que son los que se emplean para realizar el proceso, y que vamos a ver a continuación en detalle.

Supongamos la siguiente dirección Web:

```
http://librosiglibre.es/descarga
```

Si visitas esa página estás efectuando una petición a través de esa URL, la cual se compone de las siguientes partes:

- `http`: El protocolo a usar, que define la forma en que se van a comunicar cliente y servidor. Aunque este es el más habitual, existen muchos otros tales como `ftp` o `mailto`.
- `librosiglibre.es`: Esta cadena identifica la máquina donde reside la página que buscamos. Es en realidad una versión más legible para el ojo humano de un código numérico que indica la dirección concreta. El navegador lo convierte en realidad en algo como 128.118.54.228.
- `descarga`: La página que buscamos dentro de todas las que hay en esa máquina. Se expresa como una ruta a partir del directorio raíz del servidor, que no es necesariamente el directorio raíz de la maquina servidora.

El proceso mediante el que podemos ver esa página en un navegador Web comprende los cuatro pasos siguientes:

1. El cliente realiza la petición.
2. La petición se conduce a través de la red hasta el servidor.
3. El servidor busca la página y la devuelve a través de la red en caso de encontrarla, o devuelve una pagina de error en caso de no tenerla.
4. El cliente recibe la página y la representa.

[1]Técnicamente, una dirección Web como esta se conoce como URL (Uniform Resource Locator)

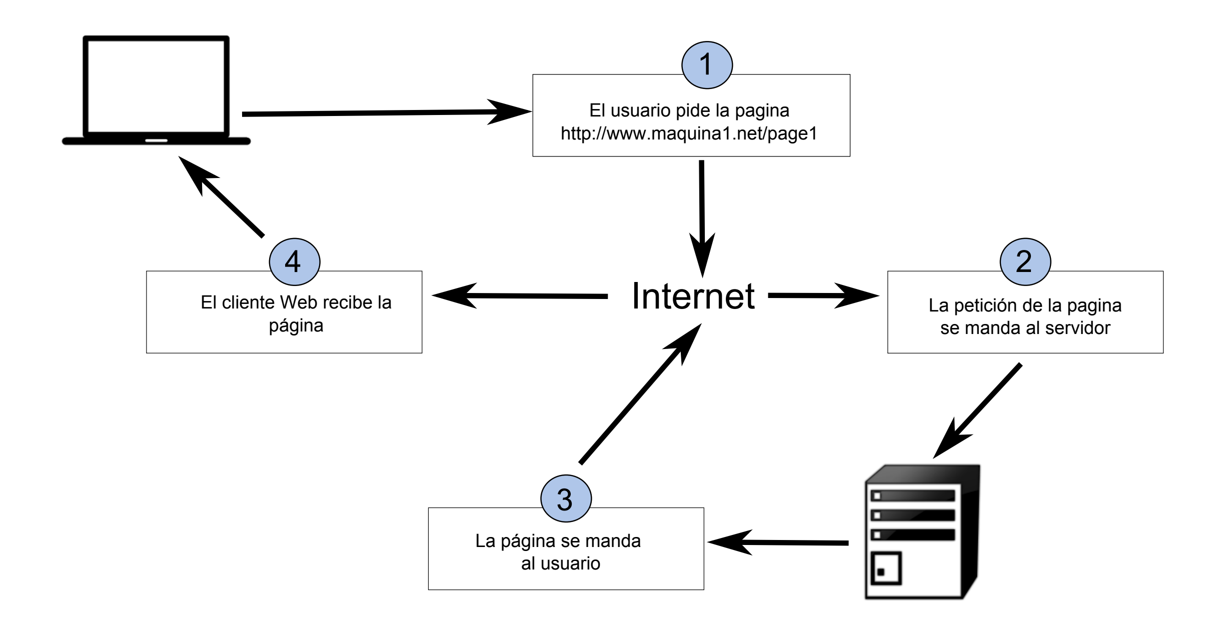


Figura 24.1: Esquema del proceso de consulta de una página Web desde un navegador.

La figura 24.1 muestra un esquema de este proceso.

24.3. El valor de las tecnologías SIG Web

Antes de abordar la parte más técnica de las tecnologías Web SIG, veamos el significado de estas y la función que cumplen. Entenderemos en este contexto como tecnologías Web SIG a todos aquellos elementos que permiten la representación de cartografía como un contenido más de una página Web. Esto es lo que se engloba bajo la denominación genérica de *Web Mapping*.

Aunque este capítulo está dedicado a las tecnologías Web dentro del ámbito SIG, y estas incluyen tanto servidores como clientes, las formas en las que se presentan los elementos del *Web Mapping* dependen fundamentalmente del cliente, el cual es en general un simple navegador.

Como vimos en el capítulo dedicado a los SIG de escritorio, estos pueden acceder a datos remotos, y para ello necesitan realizar una petición a un servidor siguiendo el esquema que hemos visto en el apartado anterior. Una vez que los datos están en el SIG (es decir, el servidor ha devuelto a este los datos que había pedido), podemos operar con ellos usando las herramientas que ya conocemos.

En un entorno Web *sensu stricto* tal como el de un navegador, las posibilidades son, no obstante, distintas, pues se trata de combinar los elementos cartográficos con los restantes elementos que forman parte habitual de una página Web. Las tecnologías Web de corte SIG se han desarrollado principalmente para su trabajo dentro de un navegador, es decir, como una alternativa a los SIG de escritorio o para alcanzar áreas nuevas en el trabajo con información geográfica digital. Su incorporación en los SIG de escritorio aumenta las capacidades de estos, pero la principal poten-

cia de estas tecnologías surge cuando se unen a otras funcionalidades de tipo Web.

En resumen, el objetivo básico que pretenden cumplir las tecnologías que vamos a ver, especialmente las del lado del cliente, es llevar las funcionalidades de un SIG a la Web, para así compartir la potencia de ambos componentes. Las ventajas de llevar el SIG a la Web en lugar de incorporar los elementos de esta última en un SIG de escritorio tradicional son notables, y existen grandes diferencias entre las soluciones que se obtienen en ambos casos. Estas diferencias tienen que ver sobre todo con los usuarios y su perfil, así como con el diseño mismo de las aplicaciones.

Mientras que un SIG de escritorio se orienta principalmente a usuarios más especializados, poder dotar a un sencillo navegador Web de capacidades de visualización o edición de información geográfica hace que estos lleguen a un público distinto y abre nuevas posibilidades. Los usuarios avanzados encuentran igualmente utilidad en el *Web Mapping*, que se complementa en muchos terrenos con los SIG de escritorio. Por su parte, los usuarios no especializados, desconocedores de otras tecnologías SIG, pueden incorporarse al ámbito SIG a través de las tecnologías Web.

Algunas de las ideas fundamentales que caracterizan a las tecnologías de *Web Mapping* y su papel actual son las siguientes:

- **No es necesario un software SIG específico**. Al menos, no es necesario desde el punto de vista del usuario, que no ha de instalar nada adicional en su ordenador. Acceder a cartografía remota e incluso a funcionalidades avanzadas basadas en esos datos no requiere más que un simple navegador Web, algo presente en cualquier ordenador hoy en día. La barrera que puede suponer el trabajar con una aplicación específica se diluye cuando incorporamos las capacidades de esta en algo tan habitual como un navegador.
- **Perfil menos técnico**. No solo las aplicaciones están pensadas para su utilización por parte de usuarios no especializados, sino que la incorporación de estos al ámbito SIG hace que la cartografía deje de ser un elemento propio de esos usuarios más técnicos. Poniendo al alcance de todos las capacidades de edición y creación de cartografía hace que cualquiera pueda generar su propia información geográfica no especializada y además ponerla a disposición de otros usuarios.
- **Potenciamiento del trabajo colaborativo**. La red es un punto de encuentro que favorece de forma natural la colaboración. Proyectos como la Wikipedia, posibles gracias a esta capacidad de Internet para facilitar el trabajo común de múltiples personas, tiene sus equivalentes en el ámbito de la información geográfica. Los SIG dejan de ser algo personal reducido al ámbito de un ordenador o una pequeña red, para ser algo global en una red de muchos SIG interconectados. Y más importante que esto, los datos también se hacen globales, pudiendo ser empleados e incluso editados por todos.
- **Información más actualizada**, incluso en tiempo real. La Web es el canal ideal para transmitir la información de forma inmediata y flexible. A las ventajas de los datos digitales sobre los analógicos en este sentido, que ya vimos en el capítulo 4, hay que sumar que la sencillez de acceso que aporta una interfaz Web hace todavía más accesible la información geográfica más reciente.
- **Independencia del sistema**. Un mapa Web puede verse y usarse del mismo modo en cualquier ordenador, con independencia del sistema operativo, el navegador e incluso el dispositivo empleado (PC, móvil, etc.). Si este mapa se basa en estándares abiertos, la solución es todavía más interoperable, como veremos en el apartado 24.7.

- **Personalización de aplicaciones**. Una de las tendencias más importantes en el ámbito del *Web Mapping* es la creación de aplicaciones que personalizan una base común para un determinado uso. Sobre una base compuesta por un juego de datos genérico (generalmente imágenes de satélite y mapas base tales como un mapa de carreteras) y una aplicación SIG, se crean pequeñas aplicaciones de forma sencilla, a las cuales se pueden añadir de modo también simple nuevos datos. Estas aplicaciones, una vez creadas, pueden a su vez incorporarse a una página Web distinta.
  Mediante una de ellas, por ejemplo, un usuario puede crear, sin excesivos conocimientos sobre SIG, una aplicación particular que ponga sobre ese juego de datos general los emplazamientos de todos aquellos que visitan su página Web. Las posibilidades en este sentido son prácticamente infinitas, y proliferan de forma exponencial en Internet.
- **Combinación de cartografía y otros elementos**. Si llevamos las capacidades SIG a un navegador, además de estas dispondremos en ese navegador de muchas otras posibilidades, tales como la representación de elementos multimedia (vídeo, sonido, etc.) o el uso de hiperenlaces. El navegador es hoy en día la aplicación versátil por excelencia, y ello hace que podamos añadir a las capacidades SIG una larga serie de otras funcionalidades no relacionadas directamente con la información geográfica, y no presentes en su mayoría en los SIG de escritorio.

La importancia de las tecnologías Web se debe, por tanto, principalmente a una razón social y no a una tecnológica, aunque es innegable que las tecnologías novedosas que se desarrollan en este campo aportan al ámbito SIG posibilidades antes desconocidas. Estas nuevas posibilidades enriquecen notablemente los SIG de escritorio si estos implementan las capacidades de acceso a datos remotos, ampliando el alcance de ese tipo de aplicaciones. Cuando se implementan, sin embargo, en un entorno puramente Web tal como en un navegador y se crea una página Web con elementos SIG, se consigue ampliar el abanico de usuarios potenciales y así también crecen las posibilidades y las formas en que el propio SIG puede presentarse.

24.4. Formas de cartografía en la Web

Las formas en las que pueden presentarse las tecnologías SIG dentro de un entorno Web varían en cuanto a su similitud con los SIG de escritorio, incorporando más o menos elementos de los que son habituales en este tipo de aplicaciones. Como parece lógico pensar, ha existido una evolución progresiva, de tal modo que en la actualidad existen más elementos propios de los SIG de escritorio dentro de las tecnologías Web SIG, y la cartografía Web hoy en día permite realizar un trabajo más similar al que se desarrolla en un SIG clásico.

Una primera y sencilla clasificación de los tipos de cartografía Web es la que divide esta en mapas *estáticos* y *dinámicos*[315].

Un mapa estático es simplemente una imagen con información cartográfica, la cual no permite ningún tipo adicional de trabajo con ella que no sea la mera observación. En este sentido, se asemeja a un mapa clásico, donde el usuario no puede interactuar directamente con el contenido del mapa. A efectos de trabajo real, las posibilidades son aún menores, ya que acciones tales como mediciones tampoco pueden realizarse, ni siquiera con medios mecánicos como el caso de un mapa en papel. Junto a esto, la resolución de una pantalla común es mucho menor que la que presenta un mapa impreso, con lo que la calidad del mapa no es comparable.

Este tipo de mapas, por tanto, no responden a las funcionalidades que un SIG ha de tener para poder prestar utilidad en el manejo y uso de información geográfica, y difieren notablemente de un SIG de escritorio, incluso en la versión más básica y primitiva de estos últimos.

Incorporar este tipo de mapas a una página Web no requiere ninguna tecnología particular, y puede llevarse a cabo con elementos genéricos tanto del lado del cliente como del servidor, pues el dato realmente no es un dato geográfico, sino una mera imagen (y esa imagen no va acompañada de información tal y como su sistema de referencia), algo para lo cual cualquier servidor o cliente actual ofrece soporte.

La figura 24.2 muestra una imagen de una primigenia cartografía Web presentada a través del visor Xerox PARC Map Viewer.

Por su parte, un mapa dinámico es aquel que no se compone de una imagen inmóvil, sino que esta varía y se adapta en función de los requerimientos del usuario o según alguna serie de parámetros prefijados. De acuerdo con esto, los mapas dinámicos pueden ser interactivos o no, dependiendo de si es el usuario quien directamente modifica la representación del mapa.

Como ejemplo de mapa dinámico no interactivo podemos citar mapas animados que encuadran una determinada zona y muestran la variación de una variable a lo largo del tiempo. Mapas de variables climatológicas o una serie animada de mapas que reflejan el avance de un incendio son ejemplos habituales de este grupo.

Tampoco en este tipo de mapas aparecen las funciones esperables en una aplicación SIG, y una vez más no se requieren tecnologías específicas para poder incorporar este tipo de elementos en una página Web.

La interactividad es la que aporta las posibilidades necesarias para comenzar a incorporar funciones SIG a la cartografía Web, y sin ella no podemos hablar en realidad de tecnologías SIG puramente dichas.

La forma de interactividad más básica que se implementa en una página Web en el trabajo con cartografía es la que permite la modificación de la manera en que los datos geográficos se visualizan. Las herramientas que permiten modificar la escala de visualización (acercarse o alejarse) y desplazar el mapa, las cuales ya nombramos como capacidades básicas en los SIG de escritorio, aportan a la cartografía Web muchas posibilidades nuevas. Entre ellas, es de destacar que mediante estas herramientas la extensión de los datos no se encuentra limitada por la propia extensión de la pantalla o la dimensión del navegador.

Si se trabaja con imágenes estáticas, trabajar con datos que cubran toda la extensión del globo implica hacerlo a una escala de muy poco detalle, pues ha de representarse toda la imagen de forma simultanea. Permitiendo que el usuario elija la escala de representación y ajuste la extensión con la que se desea trabajar, un navegador Web se convierte en una ventana hacia datos que pueden tener cualquier extensión y volumen, y hacia el trabajo con ellos de forma dinámica e interactiva.

Esto es de especial importancia si pensamos que las máquinas que se encuentran al otro lado (en el servidor) son ordenadores potentes con gran capacidad, que pueden almacenar enormes conjuntos de datos. Un conjuntos de datos con imágenes de todo el mundo a gran resolución ocupa un tamaño que probablemente lo haga inutilizable en un ordenador personal (además de que esos datos probablemente queden fuera del alcance del usuario de ese ordenador en lo que a su adquisición respecta), pero puede perfectamente almacenarse en un servidor potente, desde el que se servira en cada caso la «porción» de él que cada usuario requiere según utiliza el cliente correspondiente.

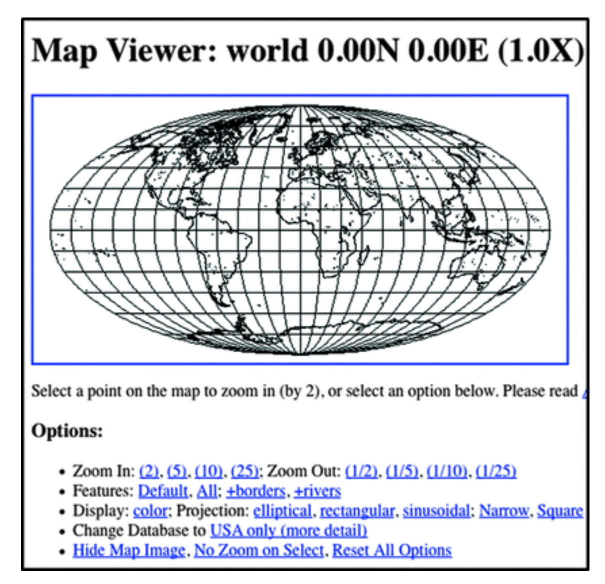


Figura 24.2: Visor de mapas Xerox PARC Map Viewer, uno de los primeros en su campo.

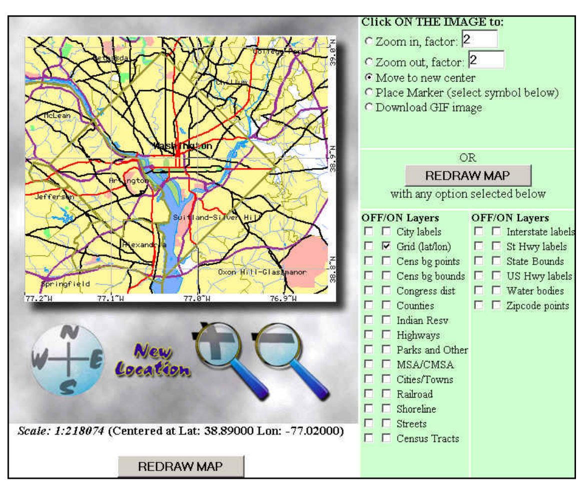


Figura 24.3: Interfaz de TIGER MapServer (año 1997).

De especial importancia para el desarrollo de estas capacidades ha sido la popularización y mejora de las tecnologías que permiten el desarrollo de las denominadas *Aplicaciones de Internet Enriquecidas* (RIA)[2]. Este tipo de aplicaciones llevan a la Web algunos elementos de las tecnologías de escritorio, y en general permiten optimizar el volumen de datos necesario para operar con la aplicación dentro del entorno del navegador.

Si no se emplean estas tecnologías, un cambio mínimo en la configuración de la pagina por parte del usuario (por ejemplo, modificar el encuadre del mapa en una aplicación SIG), requiere la recarga total de la página, de la misma forma que sucede cuando hacemos clic en un hiperenlace. En realidad, estamos pasando a una página Web distinta.

En un entorno RIA, sin embargo, se cargan al inicio (en el primer acceso a la página) los elementos que constituyen la aplicación en sí, y posteriormente se transmiten únicamente los datos que vayan siendo necesarios a medida que el usuario opere con la aplicación. Esto mejora notablemente la sensación del usuario, ya que este nunca tiene ante sí una pantalla sin contenido mientras se carga la página, puesto que esta ya no ha de cargarse de nuevo, y la carga de datos puede además realizarse mientras el propio usuario opera.

Profundizar más en estos aspectos es, no obstante, demasiado técnico para el enfoque de este libro, no siendo necesario además para la comprensión de las tecnologías Web desde el punto de vista del usuario. Tan solo es necesario diferenciar entre el comportamiento de una página Web anterior a la introducción de estas técnicas, en la cual cualquier interacción suponía una recarga completa de la página, mientras que en el

[2]Rich Internet Applications

caso de una RIA la experiencia es más fluida y cercana a la que se tiene usando una aplicación de escritorio.

La figura 24.3 muestra el aspecto de una aplicación de *Web Mapping* previa a la introducción de esta clase de tecnologías.

Además de modificar la zona representada, un usuario debe poder modificar la forma en que los datos dentro de esa zona se muestran. Es decir, debe poder cambiar el estilo de los elementos representados, variando colores o formas de la misma manera que esto puede hacerse en un SIG de escritorio. Asimismo, muchas aplicaciones Web permiten la consulta de varias capas de datos, incluso de datos provenientes de varios servidores distintos, datos que no necesariamente han de mostrarse todos simultáneamente. Igual que en un SIG de escritorio seleccionamos unas u otras capas para su visualización y podemos alterar el orden de representación de estas, también podemos realizar estas operaciones en una aplicación SIG Web.

Esto hace que una aplicación SIG dentro de un navegador se convierta en una herramienta completa para el acceso a uno o varios juegos de datos remotos cuyo contenido es abundante (no solo en extensión sino también en tipos de datos suministrados), ya que permite una gran configurabilidad y deja en manos del cliente (esto es, del usuario), la forma de tomar esos datos y mostrarlos.

Las capacidades de edición también tienen lugar en los SIG Web, ampliando las posibilidades que la interactividad más básica ofrece. Un usuario puede añadir su propia información a un SIG Web o bien modificar una capa existente empleando su navegador. Las tecnologías SIG siguen en este sentido a las tecnologías Web más generales, adoptando los conceptos de la Web 2.0 y ampliando las posibilidades de los usuarios de colaborar directamente en los contenidos de la red. Por ejemplo, *OpenStreetMap* es un sitio equivalente a la bien conocida Wikipedia, en el cual los usuarios pueden añadir sus propias descripciones de elementos geográficos que ellos mismos definen.

A estas mismas tecnologías se les puede dar usos más restringidos sin que necesariamente sea dentro de un proyecto colaborativo abierto. Por ejemplo, una administración local puede dar acceso a los propietarios de suelo para que puedan consultar su catastro, mediante un sistema de autenticación conveniente, incluso editar información de sus parcelas. Esta información puede ser de tipo no espacial (es decir, los límites de las parcelas serían fijos), ya que las capacidades de edición no han de limitarse a la componente espacial.

Por último, y aunque en la actualidad son pocos los servicios de este tipo que existen, y no pueden compararse las prestaciones con las que ofrecen los SIG de escritorio, la cartografía Web puede ofrecer herramientas de análisis. Además de representar un conjunto de datos geográficos y permitir al usuario navegar en ellos e incluso editarlos, pueden extraerse resultados a partir de esos datos.

Un tipo de aplicación bastante extendida de este tipo es el cálculo de rutas óptimas. A partir de una capa con vías de comunicación un usuario establece un punto de salida y otro de destino y la aplicación Web calcula la ruta que optimiza el tiempo empleado o la distancia total recorrida, según lo explicado en el capítulo 20.

El término *Web Mapping*, habitualmente empleado para designar a la cartografía Web, se sustituye por *Web GIS* a medida que las capacidades de las aplicaciones Web aumentan, para indicar así que todos los componentes que forman parte de un SIG en su sentido clásico, esto es, un SIG de escritorio, se incorporan a dicha aplicación Web.

24.5. Clientes y servidores

Ahora que conocemos algunas ideas generales sobre cartografía Web, veamos algo más en detalle los elementos tecnológicos que hacen posible su funcionamiento: los servidores y los clientes. Veremos en este apartado las funcionalidades que presentan y algo más de los fundamentos tecnológicos en los que se basan, que se apoyan sobre las ideas básicas de funcionamiento de Internet que ya vimos anteriormente.

En el sistema cliente-servidor se presentan las siguientes características principales:

- **El servidor brinda servicio a múltiples clientes**. Los clientes, por su parte, también pueden acceder a servicios en varios servidores, aunque esa multiplicidad es mucho más relevante en el caso del servidor. Piénsese, por ejemplo, en un navegador Web con el que podemos acceder a varias páginas y un servidor de una de dichas páginas. Mientras que en el cliente no accedemos simultáneamente a un gran número de páginas (si la pagina es estática solo usamos el servicio al cargarla, y no cargamos más de una capa en un instante dado), el servidor debe estar preparado para responder a muchas peticiones simultaneas y satisfacer la demanda de muchos clientes en un instante concreto.
- **Los clientes no dependen de la ubicación física del usuario**, el sistema operativo o la arquitectura física de la máquina. Esto es así porque el cliente no necesita conocer la lógica interna del servidor para usar sus servicios. Lo único necesario es que el servidor pueda exponer una interfaz externa que actúe como un modo de comunicación para recibir las peticiones del cliente, siendo esta comunicación siempre transparente para este último.
- **La carga de proceso se puede repartir entre cliente y servidor**. En función del servicio y de las capacidades del cliente, el trabajo puede dividirse de una u otra forma entre las partes implicadas.

24.5.1. Servidores

El servidor es el elemento encargado de ofrecer el servicio como tal, respondiendo a las peticiones del cliente. A medida que los clientes se hacen más complejos y presentan mayor número de funcionalidades, también los servidores deben ser capaces de proporcionar servicios más elaborados. Las capacidades fundamentales a las que responden los servidores dentro del ámbito SIG pueden dividirse en los siguientes grupos:

- **Servir representaciones de los datos**. Los servicios de cartografía Web, tanto en sus orígenes como en la actualidad, son eminentemente gráficos, y en última instancia lo que la aplicación Web correspondiente va a hacer es mostrarnos algún tipo de imagen con un mapa formado a partir de una serie de datos geográficos.
  El servidor puede responder directamente a este tipo de necesidades, preparando una imagen a partir de los datos geográficos de los que dispone. En el caso de que estos sean ya imágenes —por ejemplo, imágenes de satélite u ortofotos—, bastará servir estas, transmitiendo una versión escalada de las dimensiones exactas que el cliente necesite para representar en pantalla. En caso de que los datos sean de tipo vectorial, o bien ráster sin una forma de representación implícita —por ejemplo, un Modelo Digital del Terreno— es necesario emplear algún método para asignarles dicha representación. Este puede ser asignado por defecto por el servidor, que establecerá una simbología fija, o bien ofrecer un servicio más complejo en el que el cliente no solo pide una representación gráfica de una serie de datos para una zona dada, sino que además puede especificar *cómo* crear esa representación.

Asimismo, el servidor puede ofrecer la posibilidad de seleccionar los datos empleados para crear la representación gráfica. En términos de un SIG de escritorio esto es equivalente a seleccionar qué capas se van a representar de entre el total de las que se encuentran abiertas o bien en nuestro catálogo de datos al que tenemos acceso desde el SIG. En el caso de un servicio Web, el servidor dispone de una serie de capas a las que puede acceder, y a la hora de servir una imagen puede preparar esta usando unas u otras según las necesidades que el cliente especifique a la hora de hacer la petición del servicio. De igual modo, el orden en que se desea que las capas se pinten en el mapa también debe poder ser especificado por el cliente.

- **Servir los datos directamente**. Una opción más flexible que lo anterior es que el servidor provea directamente los datos geográficos y sea después el cliente quien los utilice como corresponda, bien sea simplemente representándolos —en cuyo caso debería ser el propio cliente quien establezca la simbología, ya que esta tarea ya no queda en manos del servidor— o bien trabajando con ellos de cualquier otra forma, como por ejemplo analizándolos.

Aunque las posibilidades son mayores en este caso, se requieren por parte del cliente unas capacidades mayores, ya que mientras que representar una imagen es algo sumamente sencillo desde el punto de vista técnico, crear esta a partir de los datos geográficos es más complejo.

- **Servir consultas**. Un paso más allá en la funcionalidad que puede ofrecer el servidor es responder a *preguntas* realizadas por el cliente relativas a los datos, ya sean estas relativas a la parte espacial de dichos datos, o bien a su componente temática. El servidor puede ofrecer como respuesta conjuntos reducidos de los datos de los que dispone, o valores que describan a estos. Estas consultas pueden ser útiles, por ejemplo, para establecer filtros previos cuando se dispone de un conjunto amplio de orígenes de datos.

Un cliente Web puede obtener datos de distintos servidores, y puede consultar si, para un zona dada, estos servidores disponen de información, sin más que consultar la extensión cubierta por los datos de cada uno de ellos y comprobar si se interseca con la región de interés. En función de la respuesta, puede o no realizarse posteriormente el acceso a los datos en sí. Como ya vimos en el capítulo 9, los *metadatos* son de gran utilidad para conseguir que este tipo de consultas se realicen de forma eficiente.

- **Servir procesos**. Por último, un servidor puede ofrecer nuevos datos, espaciales o no espaciales, resultantes de algún tipo de proceso o cálculo a partir de datos espaciales. En este caso, el proceso constituye en sí el servicio ofrecido por el servidor, y el cliente debe definir los parámetros de entrada de este y los posibles parámetros de ajuste que resulten necesarios. Los datos con los que se trabaja pueden ser proporcionados por el cliente, incorporándolos a su propia petición, o bien pueden residir en el propio servidor. En este último caso, el servidor ofrece tanto los datos, como la posibilidad de extraer resultados a partir de ellos, es decir, los datos y una herramienta para explotarlos. También pueden emplearse datos en un servidor distinto, a los que el servidor de procesos puede acceder si estos están disponibles, convirtiéndose en cliente de ese segundo servidor (Figura 24.4).

Las posibilidades que estos servicios brindan son muy numerosas. Por una parte, pueden añadirse funcionalidades avanzadas a interfaces Web, llevando a estas las capacidades propias de los SIG de escritorio. Por otra, la difusión de algoritmos de análisis geográfico resulta más sencilla, pudiendo ofrecerse estos a todo tipo de usuarios sin necesidad de ningún software especializado. Y por último, en ciertos casos pueden rebajarse

los tiempos de proceso, ya que, en el caso de operaciones complejas, la mayor potencia del servidor respecto al cliente puede resultar en un mayor rendimiento. El reparto de tareas entre varios servidores (computación distribuida) es otra de las posibilidades que pueden a su vez ampliar la eficiencia de los procesos.

24.5.2. Clientes

El cliente es el elemento que utiliza los datos proporcionados por el servicio. Para ello, realiza una petición a la que el servicio responde enviando dichos datos, que serán los que después se emplearán para realizar cualquier otra tarea, principalmente la representación de estos para que el usuario pueda visualizarlos. El cliente es, de este modo, el intermediario entre el usuario y los servicios y datos que el servidor ofrece.

Como hemos visto al estudiar los servidores, las principales capacidades de estos implican la transmisión de imágenes con cartografía ya elaborada, o bien directamente capas, ya sean de tipo ráster o vectoriales. En algunos casos, el servicio ofrecido es un servicio de procesos, pero su resultado generalmente es también una capa, por lo que, desde el punto de vista del cliente, la funcionalidad es en cierto modo similar (aunque internamente requiera una implementación por completo distinta).

El cliente, por tanto, debe disponer de capacidades para formular peticiones a servidores como los anteriormente descritos, así como para emplear las posibles respuestas que estos devolverán. Estas últimas capacidades incluyen por lo general componentes de representación, habitualmente con la forma típica de un visor en el que se permite cambiar la escala y desplazar la vista, tal y como ya vimos en el capítulo 23. No obstante, pueden variar de un cliente a otro, desde el mínimo necesario para simplemente representar los datos obtenidos del servidor hasta conjuntos de funcionalidades mucho más avanzadas pensadas para un uso intensivo de esos mismos datos.

Distinguimos así dos tipos de clientes en función de las capacidades que tengan: *clientes ligeros* y *clientes pesados*.

Un cliente ligero es un programa de tamaño más o menos reducido, lo cual va consecuentemente asociado a unas capacidades limitadas. Hablamos de clientes ligeros cuando nos referimos a *Web Mapping* y a clientes que se ejecutan sobre un navegador Web, los cuales son siempre sencillos en cuanto a sus funcionalidades. En el momento de la carga de la página Web que contiene al cliente, el navegador descarga toda la lógica del programa, lo cual hace necesario limitar el tamaño de este.

No obstante, los clientes Web empiezan progresivamente a ampliar sus posibilidades, y en ello juegan un importante papel otros servicios distintos a los de mapas o los de datos, como pueden ser los de procesos. Estos permiten que las funcionalidades adicionales no se implementen en el propio cliente (y por tanto sin aumentar en exceso su tamaño y sin disminuir su «ligereza»), sino que se accede a estas también como servicios remotos.

La evolución de la cartografía Web en esta dirección se dirige desde el *Web Mapping* al *Web GIS*, tal y como comentamos algunas páginas atrás.

A diferencia del cliente ligero, el cliente *pesado* es una aplicación individual que no se ejecuta sobre otra aplicación soporte como puede ser un navegador Web. Al ser un programa independiente, debe ocuparse de toda la lógica del proceso y de proveer todas las funcionalidades necesarias, por lo que su tamaño es generalmente mayor.

Pese a ello, un cliente pesado no ha de ser necesariamente más potente y con más funcionalidades que uno ligero (aunque habitualmente lo

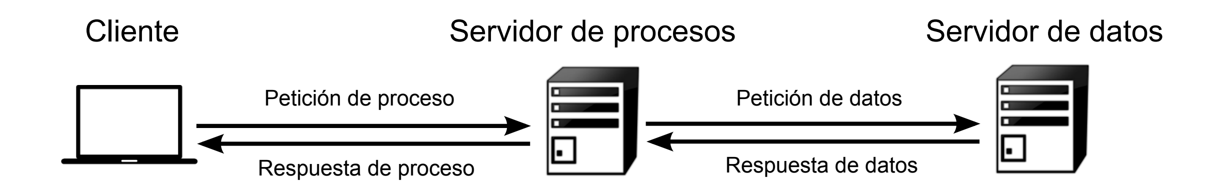


Figura 24.4: Esquema de acceso a un servicio de procesos remotos, el cual a su vez utiliza datos de un segundo servidor. El encadenamiento de procesos permite ampliar notablemente la utilidad de estos.

es), ya que existen aplicaciones muy sencillas con capacidad para conectarse a servicios de mapas, que ofrecen poco más que un visor de cartografía. La diferencia no estriba en las capacidades del programa, sino en el enfoque a la hora de implementar este y el uso o no de otra aplicación «plataforma», generalmente en forma de un navegador Web.

Los clientes pesados suelen permitir el uso de datos no procedentes directamente del acceso a servicios, tales como datos en ficheros locales, y no están pensados exclusivamente como clientes, sino como aplicaciones más amplias que además disponen de capacidades para aprovechar un determinado tipo de servicios.

Dicho de otro modo, un cliente pesado tal y como un SIG de escritorio tiene utilidad aunque no se emplee como cliente de ningún servicio y no se disponga de conexión a red alguna, ya que puede alimentarse con datos locales, y todas sus restantes funcionalidades (análisis, preparación de cartografía, etc.) pueden aprovecharse con dichos datos.

24.6. Limitaciones y problemas de la cartografía Web

Trasladar las ideas de los SIG de escritorio a la Web no es sencillo, por cuanto el entorno en el que nos movemos es muy distinto en uno u otro caso. La Web tiene sus propias limitaciones e inconvenientes, que en muchos casos no existen en el caso de una aplicación de escritorio, y este hecho presenta dificultades complejas de salvar, obligando a desarrollar soluciones alternativas.

Una limitación básica es la impuesta por el propio navegador como marco de trabajo. Las propias ventajas que este aporta son también responsables de ciertas limitaciones, ya que en el desarrollo de una aplicación SIG Web no se tiene la misma libertad que al desarrollar una aplicación de escritorio. Este no es un problema exclusivo del *Web Mapping*, sino en general de todas las aplicaciones Web. Pese a los avances que han tenido lugar en este sentido y la rápido evolución de las tecnologías Web, siguen sin poder ofrecer exactamente las mismas funcionalidades en lo que a interfaces respecta.

A lo anterior debemos sumar el hecho de que las tecnologías Web en general son recientes y en cierto modo inmaduras, y aunque se emplea gran cantidad de medios y esfuerzo en el ámbito Web debido a su vital importancia en la actualidad, una buena parte de los elementos tecnológicos sobre los que se fundamenta el *Web Mapping* actual no están todavía completamente desarrollados y necesitan aún evolucionar.

El aspecto más problemático es, no obstante, la propia red, especialmente en lo que respecta a su fiabilidad y rendimiento. Todos los datos que el cliente emplea en una aplicación de cartografía Web provienen de la red, y por tanto existe

una fuerte dependencia entre la aplicación y el funcionamiento tanto de esta como del servidor que a través de ella nos proporciona esos datos.

Si abrimos un archivo con datos espaciales en nuestro ordenador desde un SIG de escritorio, podemos casi garantizar que esa misma operación funcionará de igual modo si la repetimos en otro momento. Tener esa misma seguridad cuando se trabaja con datos remotos no es tan sencillo, ya que la red puede no funcionar o el servidor puede estar recibiendo en este momento gran cantidad de peticiones de otros clientes y no ser capaz de gestionarlas eficientemente y ofrecernos al instante respuesta a nuestra petición. En definitiva, las mismas circunstancias que afectan a todas las aplicaciones Web y que son conocidas por todos.

El rendimiento de la red es más importante aún si cabe en el caso de trabajar con información geográfica, ya que los datos suelen ser voluminosos. Visualizar un mapa y que este pueda desplazarse y modificarse de forma igual de fluida que al trabajar con una aplicación de escritorio requiere por un lado un ancho de banda suficiente para transmitir la gran cantidad de datos necesarios, y por otro la implementación de algunas técnicas particulares que facilitan este proceso. Por su importancia, veremos en detalle las técnicas de *tiling* (división horizontal de los datos geográficos en teselas) y *cacheo* (almacenamiento temporal de datos en la máquina del cliente), utilizadas habitualmente en la actualidad.

24.6.1. *Tiling* y *cacheo*

Dos técnicas básicas que se emplean actualmente en los clientes Web que manejan información geográfica son el *tiling* y el *cacheo*. Estas técnicas permiten que la experiencia de trabajar con información geográfica dentro de una aplicación SIG Web sea más agradable, logrando una mayor fluidez y superando en cierta medida las limitaciones de la red. Aunque es cierto que cada vez disfrutamos de mayores anchos de banda y velocidades de transmisión más altas, también aumentan de igual modo los volúmenes de datos manejados, con lo que las dificultades siguen existiendo de manera similar.

El *tiling* es una técnica consistente en dividir las imágenes con las que se trabaja en imágenes menores que formen un mosaico. Esto permite un trabajo más rápido, al utilizar unidades mínimas de menor tamaño y poder reducir la necesidad de transmitir datos a través de la red si se realiza una gestión correcta del conjunto de elementos de ese mosaico.

Esta división es similar en forma a la propia que se da en los datos originales, ya que, como sabemos (véase sección 4.4), estos también se encuentran divididos horizontalmente. No obstante, se trata de una estrategia propia del sistema cliente-servidor, que divide las propias imágenes que luego se representarán en este último, de forma que en lugar de transmitir una única imagen se transmiten varias de menor tamaño y la información correspondiente a la posición relativa de estas.

El *cacheo*, por su parte, es una técnica no exclusiva del ámbito SIG, sino de la Web en general, y consiste en almacenar de forma temporal los datos obtenidos de un servidor en la máquina local o bien en una máquina intermedia (*proxy*). De este modo, si volviera a resultar necesario acceder a esos datos, no han de pedirse al servidor, sino que pueden recuperarse de la copia local, con las ventajas que ello tiene en cuanto a la velocidad de acceso y la fiabilidad del proceso.

El uso conjunto de *tiling* y *cacheo* puede disminuir sensiblemente el volumen de datos a transmitir para, por ejemplo, modificar el encuadre de un mapa en una aplicación SIG Web. La figura 24.5 muestra un ejemplo sencillo que servirá para

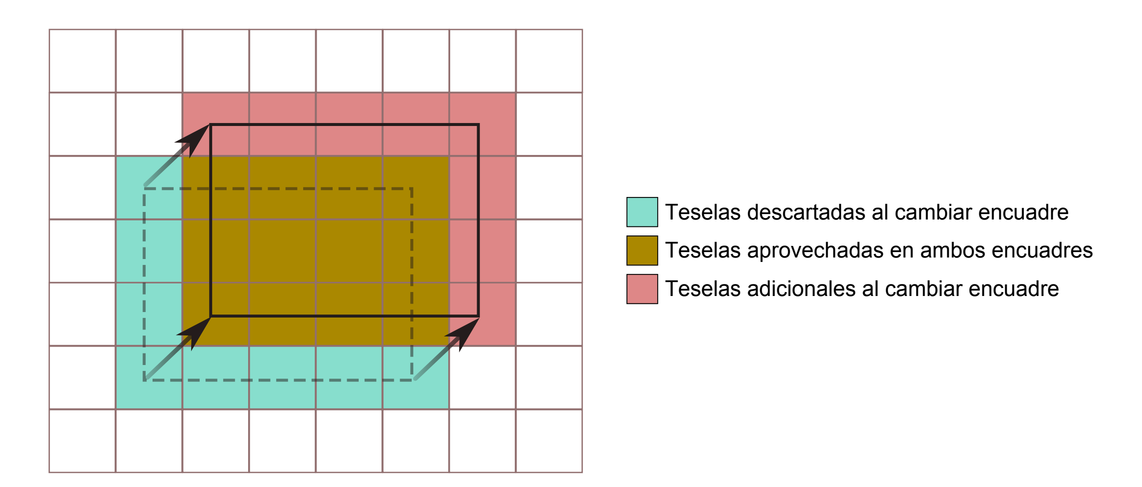


Figura 24.5: Esquema del uso de *tiling* y *cacheo* para optimizar la transmisión de datos en una aplicación SIG Web

comprender el ahorro de datos que puede conseguirse con el uso conjunto de estas técnicas.

En la figura puede verse el dato global al que se accede, dividido en una serie de unidades. Ello no quiere decir que el dato original tenga ese número de divisiones o que existan otros tantos ficheros. Puede tratarse de un único fichero, o de un número muy elevado de ellos. Las divisiones se realizan a efectos de crear el mosaico de imágenes a la hora de transmitir estas.

Inicialmente, la aplicación Web encuadra una región que cubre 20 elementos o teselas. Si el usuario desplaza el encuadre para que cubra otro área distinta, como en el caso mostrado en la figura, el cliente realizará una nueva petición y obtendrá una nueva imagen, que tendrá exactamente el tamaño con que esa imagen va a representarse. Este es exactamente el mismo tamaño que la imagen que encontramos inicialmente en el encuadre original, y por tanto la representación de este encuadre original y posteriormente el encuadre modificado requiere transmitir dos imágenes que cubren cada una de ellas veinte teselas.

Si, por el contrario, aplicamos conjuntamente las técnicas anteriores de *tiling* y *cacheo*, al variar el encuadre no es necesario obtener del servidor una imagen que cubra todo el área a representar, sino tan solo los 8 elementos correspondientes a la zona no cubierta por la imagen inicial, ya que los restantes ya habrán sido obtenidos con anterioridad y se encontrarán almacenados (*cacheados*) en nuestro ordenador. Es decir, el cliente crea la imagen a representar con 8 subimágenes pedidas al servidor y otras 12 ya descargadas previamente, reduciendo sensiblemente el volumen de datos pedidos al servidor.

Cuando este esquema de funcionamiento se combina con otras tecnologías Web que añaden a su vez mayor fluidez y una mejor respuesta de la aplicación, el resultado es una aplicación SIG altamente funcional, cuyo comportamiento se asemeja en cuanto a rendimiento al de un SIG de escritorio trabajando con datos locales.

Este tipo de técnicas no son exclusivas de los SIG en Internet, sino que también se aplican por igual al caso de SIG de escritorio cuando estos actúan como clientes y acceden a datos remotos.

La combinación de *tiling* y *cacheo* se lleva a cabo a múltiples escalas, de forma que se reduce el número de operaciones a realizar y se obtiene un mayor rendimiento. Se emplean las denominadas *pirámides*, que ya vimos en el apartado 3.5.2 dedicado a la generalización cartográfica en un SIG. Estas pueden ser empleadas también en el lado del servidor, incluso cuando este sirve mapas creados a partir de cartografía vectorial. Para evitar tener que rasterizar los datos vectoriales cada vez que se realiza una petición (lo cual supondría un gran coste en términos de proceso), se rasterizan de antemano a distintas escalas, y se *cachean* en el servidor. De esta forma, cuando el cliente efectúa la petición, ya se dispone de una imagen que servirle, sea cual sea la escala que pida..

Una técnica de reciente aparición es la denominada *tiling vectorial*. Aplicando los mismos principios que el *tiling*, es decir, la subdivision de los datos de forma regular, las capas vectoriales se «trocean» en el origen y se envían después solamente los datos necesarios para el área cubierta en el cliente. Combinando este enfoque con el uso de capas con distinto detalle según la escala, se logran dos ventajas:

- Disminución del volumen de datos.
- Capacidad de modificar la simbología en el cliente.

Al enviar los datos en lugar de una representación de estos, el cliente es quien debe establecer la simbología, lo cual permite que sea el usuario quien seleccione cómo representar los elementos vectoriales. Al mismo tiempo, se logran ventajas en la experiencia de usuario, debidas principalmente a la escalabilidad de los datos vectoriales, que permite por ejemplo presentar transiciones más fluidas cuando se modifica la escala del mapa.

24.7. Estándares

Para garantizar el buen funcionamiento de un sistema cliente-servidor, es importante definir de forma adecuada cómo se establece la comunicación entre clientes y servidores, de forma que estos primeros no solo puedan obtener los propios datos geográficos de estos últimos, sino también realizar consultas o conocer qué otras funcionalidades se encuentran disponibles.

En otras palabras, resulta necesario definir una *lingua franca* para que todas las comunicaciones se produzcan de forma fluida. Esto obliga a establecer una cierta normalización y crear elementos estandarizados que sean conocidos e implementados por las distintas partes, y hacerlo para cada uno de los servicios ofrecidos, así como para los propios datos. Esta *lingua franca* es lo que denominamos un *estándar*.

El modelo de cliente-servidor en términos tecnológicos no es muy diferente de la idea de un cliente y un proveedor de servicios en la vida real. Una persona (el cliente) que quiera adquirir un producto de un distribuidor (el servidor) debe igualmente comunicarse con él para preguntarle si dispone del producto deseado, realizar una petición de este y después recibirlo cuando el distribuidor se lo envíe.

Imaginemos ahora la situación en la que una persona en España desea adquirir un producto electrónico de un proveedor chino. En primer lugar, es probable que tenga dificultades para entender el catálogo de productos, pues este describirá cada uno de ellos en chino. Si consigue localizarlo y desea adquirirlo, es igualmente probable que encuentre dificultades para comunicárselo al proveedor, ya que seguirá existiendo la misma barrera lingüística. Y si finalmente recibe el producto, puede tener dificultades al utilizarlo, ya que este puede funcionar a un voltaje distinto al

de la red eléctrica española o bien estar preparado para un tipo de enchufe distinto.

Este pequeño ejemplo nos hace ver que en la relación cliente-servidor pueden surgir problemas derivados de la falta de elementos comunes entre ambos actores. Si todos los elementos que toman parte en el establecimiento de esa relación comercial estuvieran normalizados y fueran únicos, un comprador de cualquier parte del mundo podría de forma inmediata comprar un dispositivo a cualquier vendedor de otro país comunicándose en un único idioma, y tener después la garantía de poder usarlo sin problemas.

En el ámbito de la información geográfica la situación es similar a la anterior. Hay muchos formatos distintos para almacenarla y muchas formas distintas de transmitirla, y ello dificulta el trabajo. Igual que los clientes españoles no hablan el mismo idioma que los vendedores chinos, no todos los clientes SIG hablan el mismo idioma que todos los servidores, y dos cualesquiera de ellos no han de «entenderse» necesariamente.

De hecho, históricamente los distintos fabricantes de clientes definían por sí mismos la forma en que sus programas se comunicaban, que no coincidía con la del resto de fabricantes. Un cliente de un fabricante dado no podría acceder a los servicios de un servidor creado por un fabricante distinto. Este paradigma, característico del software privativo, es un problema, pues dificulta el acceso a los datos.

En circunstancias ideales, debe existir una total *interoperabilidad* con independencia de los formatos y las aplicaciones empleadas, pudiendo interactuar entre sí los distintos clientes y servidores. Los estándares son el elemento que va a permitir esa interoperabilidad, definiendo el marco común que clientes y servidores emplearán para entenderse. En un contexto altamente heterogéneo tanto en datos como en herramientas, lograr esto no resulta una tarea sencilla[316], y los estándares son los encargados de aportar homogeneidad tecnológica.

24.7.1. Estándares abiertos e interoperabilidad

La interoperabilidad implica que podemos sustituir unos elementos del sistema en el que se incluyen los clientes y servidores por otros distintos, teniendo la seguridad de que van a interaccionar entre ellos sin dificultades. Las funcionalidades que un cliente o servidor nos ofrece pueden ser distintas a las de otro, pero independientemente de su origen (independientemente del fabricante), si esos elementos implementan un estándar dado, siempre podrán interactuar con todos aquellos que también lo implementen.

La clave, por tanto, está en los estándares, y en particular en que estos sean estándares *abiertos*. Un estándar es un documento o práctica que busca armonizar los aspectos técnicos de un producto o servicio.

Un estándar se considera como tal cuando es empleado por un grupo o comunidad, que lo acepta para la definición de las características de ese producto o servicio en su seno. Si únicamente es el uso del estándar el que lo ratifica como tal, se denomina estándar *de facto*. El formato *shapefile* para capas vectoriales, por ejemplo, es uno de estos estándares, ya que está ampliamente difundido y existe tal cantidad de datos en este formato que todas las aplicaciones deben implementarlo para tener valor práctico.

Existen estándares que se convierten en normas o estándares *de iure*, cuando estos son promovidos por algún organismo oficial de normalización o su uso se impone con carácter legal.

Un estándar *abierto* es aquel cuya definición se encuentra disponible y todo aquel que lo desee puede conocerla y emplearla para el desarrollo de la actividad relacionada con ese estándar. En

nuestro campo de trabajo, eso quiere decir que cualquier desarrollador que desee crear un nuevo cliente o servidor para datos SIG puede acceder al estándar y desarrollar en base a este.

Los principios fundamentales de los estándares abiertos son los siguientes:

- **Disponibilidad**. Los estándares abiertos están disponibles para todos el mundo para su lectura y uso en cualquier implementación.
- **Máxima posibilidad de elección para los usuarios finales**. Los estándares abiertos crean un mercado competitivo y justo, y no bloquean a los usuarios en el entorno de un vendedor particular. Desde el punto de vista de quienes venden la tecnología SIG, esto no es tan ventajoso, ya que permite la aparición de competidores que antes no podían existir. Si un fabricante basa sus productos en un estándar cerrado definido por él mismo, otros no pueden elaborar soluciones que trabajen con esos productos, ya que no conocen el estándar empleado.
  Asimismo, el fabricante puede, por ejemplo, cambiar el estándar utilizado por su producto de servidor, y obligar a los consumidores y a todo aquel que quiera utilizar un servicio basado en ese servidor a que actualicen también los clientes, pues los anteriores ya no podrán comunicarse con el nuevo servidor. Utilizando estándares abiertos, la competencia entre fabricantes ha de basarse puramente en las capacidades que ofrecen, con lo que los consumidores ganan en calidad de los productos y en posibilidades de elección.
- **Gratuidad**. Implementar un estándar es gratuito, sin necesidad de pagar, como en el caso de una patente. Los organismos que generan los estándares pueden cobrar una cierta cantidad por acceder a la definición de los estándares, con objeto de financiar así la labor que desarrollan, y también pueden cobrar por emitir certificados de que un determinado producto o servicio se ha desarrollado de acuerdo con el estándar.
- **Discriminación**. Los estándares abiertos y las organizaciones que los desarrollan no favorecen de ningún modo a uno u otro implementador sobre los restantes.
- **Extensión o creación de subconjuntos de un estándar**. Los estándares abiertos pueden extenderse o bien presentarse como subconjuntos del estándar original.
- **Prácticas predatorias**. Los estándares abiertos pueden tener licencias que requieran a todo aquel que desarrolle una extensión de dicho estándar la publicación de información acerca de esa extensión, y el establecimiento de una licencia dada para todos aquellos que creen, distribuyan y vendan *software* compatible con ella. Un estándar abierto no puede prohibir de otro modo el desarrollo de extensiones.

Para tener una noción de lo que en la práctica realmente significa el uso de estándares abiertos en el campo de los SIG, podemos ver la figura 24.6, donde se representa el esquema de una arquitectura no interoperable. Es decir, una arquitectura que no se basa en este tipo de estándares.

Los datos que se encuentran en cada base de datos son accesibles únicamente a través de un único cliente, que es aquel correspondiente al servidor que ofrece servicios basados en esos datos. Los restantes datos quedan fuera del alcance de ese cliente, ya que no es capaz de acceder a ellos. Las diferentes soluciones cliente-servidor crean en esta situación un conjunto de islas tecnológicas, cada una completamente independiente y sin posibilidad alguna de interactuar con las restantes.

Entre los principales inconvenientes de una arquitectura no interoperable como la representada podemos citar los siguientes:

- **Desperdicio de recursos**. Cada servicio debe gestionar sus propio conjunto de datos, lo cual requiere abundantes recursos y no es sencillo, además de implicar un elevado coste.

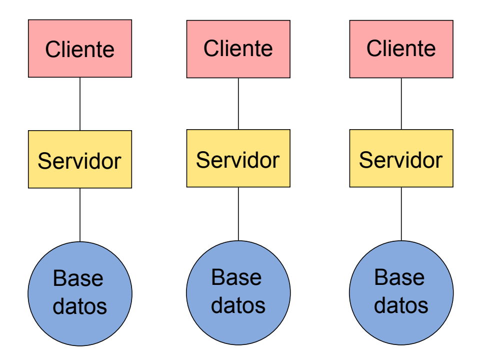


Figura 24.6: Esquema de una arquitectura no interoperable.

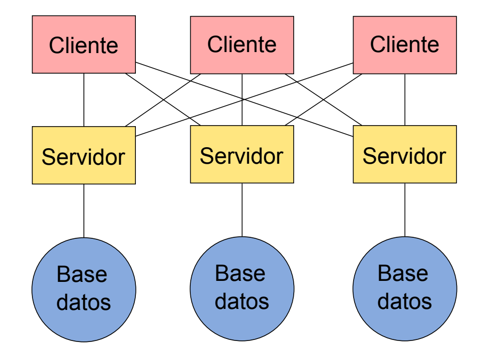


Figura 24.7: Esquema de una arquitectura interoperable.

- **Necesidad de conocer múltiples clientes**. Si para acceder a cada servicio necesitamos su cliente particular, acceder al conjunto de servicios ofrecidos por esos servidores requiere por parte de los usuarios aprender a utilizar tantos clientes como servidores existan.
- **Imposibilidad de combinar datos**. Dos datos a los que pueda accederse a través de dos servidores distintos no podrán utilizarse simultáneamente en un único cliente, ya que este no podrá comunicarse con ambos servidores. Un análisis que requiera distintos tipos de datos no podrá realizarse si todos ellos no se ofrecen a través de un mismo servidor.
- **Imposibilidad de combinar funcionalidades**. Los datos ofrecidos por un servidor pueden usarse para el desarrollo de muchas tareas. Estas tareas requieren que las correspondientes herramientas estén disponibles en los clientes, y estos se diferencian notablemente, de la misma forma que lo hacen también los SIG de escritorio entre sí. Si acceder a los datos a través de un servidor solo se puede hacer empleando un cliente concreto, no existe la posibilidad de aprovechar las funcionalidades de otro cliente sobre esos mismos datos, y el usuario ve así limitadas sus posibilidades de trabajo.

En contraste con lo anterior, tenemos una situación de plena interoperabilidad basada en estándares abiertos como la representada en el esquema de la figura 24.7.

En este caso, existe un servidor que es el que gestiona y ofrece los servicios para cada base de datos, pero a él pueden acceder todos los clientes, ya que por el hecho de estar basados en estándares abiertos es posible una comunicación plena entre dos cualesquiera de ellos.

En esta situación, un usuario puede emplear su cliente favorito (siempre que este implemente los estándares pertinentes) para acceder a muchos servicios distintos, o bien puede utilizar varios clientes para acceder a unos mismos datos, eligiendo en cada momento el que más le convenga según sus necesidades. Las posibilidades de trabajo se multiplican cuando la arquitectura del sistema se fundamenta en estándares abiertos.

Las ventajas no son solo para los usuarios, sino también para los desarrolladores. A la hora de crear un cliente, no es necesario comprobar que este se comunica bien con todos los servidores y funciona correctamente, sino simplemente seguir las especificaciones del estándar. Todo aquel servidor que las implemente funcionará sin dificultades, ya que el estándar garantiza la buena comunicación y la interoperabilidad.

24.7.2. Entidades creadoras de estándares

Crear un estándar no es una labor sencilla. Se han de recoger las principales necesidades y armonizar todas ellas en una especificación única, de modo que clientes y servidores que implementen ese estándar sean de la mayor utilidad posible para todos los usuarios.

Existen organizaciones dedicadas a redactar las especificaciones correspondientes a estándares que cubran los distintos servicios, así como a promoverlas y mejorarlas. Los estándares más habituales en el campo de la información geográfica son elaborados por tres organizaciones: el Open Geospatial Consortium (OGC), ISO y W3C.

Open Geospatial Consortium (OGC)

El Open Geospatial Consortium es una organización internacional y voluntaria dedicada a la elaboración de estándares. En el OGC participan más de 350 organizaciones miembro, incluyendo entre ellas a los principales fabricantes del sector, agencias nacionales, grupos de investigación u organizaciones sin ánimo de lucro, entre otros. Estas organizaciones miembro colaboran para alcanzar consensos y desarrollar e implementar estándares en el ámbito de los contenidos geoespaciales.

Algunos de los estándares OGC más relevantes, los cuales veremos a lo largo de este capítulo, son los siguientes:

- **WMS**. Para obtener imágenes de mapas.
- **WCS**. Para obtener y consultar coberturas.
- **WFS**. Para obtener y editar entidades geográficas y sus atributos asociados.
- **WPS**. Para servicios de procesos remotos.
- **GML**. Para almacenamiento de información geográfica.
- **CSW**. Para consultas en catálogos.

Cada uno de estos estándares está descrito en una especificación, y estas están sujetas a cambios y mejoras, existiendo varias versiones en cada caso.

ISO

ISO es una organización internacional dedicada a la elaboración de estándares no solo en el ámbito geográfico, sino en todas las áreas. ISO es responsable, por ejemplo, de estándares bien conocidos y aplicados en la industria actual, tales como los relacionados con la gestión medioambiental en empresas o los estándares de calidad.

Dentro de ISO existen diversos comités técnicos, cada uno de los cuales se encarga de definir los estándares correspondientes a un campo de trabajo. El comité ISO/TC 211 es el responsable de aquellos relacionados con la información geográfica digital.

ISO redacta Especificaciones Técnicas y Estándares Internacionales, catalogando estos con un número que los identifica. Los elaborados por ISO/TC 211 corresponde a la serie 19100.

Existe una estrecha relación entre ISO y OGC, y los estándares elaborados por ambas organizaciones son muchos de ellos muy similares o incluso idénticos. De hecho, algunos de los estándares desarrollados por el OGC, como WMS o GML, citados anteriormente y que en breve detallaremos, son también estándares ISO.

W3C

El Consorcio World Wide Web (W3C) es un consorcio internacional donde las organizaciones miembro, personal a tiempo completo y el público en general, trabajan conjuntamente para desarrollar estándares Web. Según su propia definición, la misión del W3C es «guiar la Web hacia su máximo potencial a través del desarrollo de

protocolos y pautas que aseguren el crecimiento futuro de la Web».

El W3C no guarda una relación directa con los SIG, pero parece lógico pensar que todo aquello que se haga en el seno de Internet debería acomodarse a las pautas establecidas por este consorcio, en especial si lo que se desea es maximizar la interoperabilidad, como ya hemos visto que resulta de interés en el ámbito SIG. Puesto que la mayoría de los estándares abiertos que vamos a ver en este capítulo se aplican sobre tecnologías que operan en la red, estos se han de fundamentar siempre que sea posible en otros existentes desarrollados por el W3C, o al menos seguir las recomendaciones de este organismo.

Visto de otro modo, el W3C persigue objetivos similares a los de las organizaciones que elaboran estándares para la información geoespacial, pero su campo de actuación es la red en términos generales.

De entre todos los elementos definidos por el W3C, resulta de especial importancia el lenguaje XML (eXtensible Markup Language[3]). XML no es un lenguaje en sí, sino que permite definir la gramática de otros lenguajes. Es lo que se conoce como *metalenguaje*. De este modo, puede utilizarse para definir reglas para crear formas de expresión que permitan recoger cualquier tipo de información. Esto hace que pueda emplearse para el intercambio de información de toda clase, y es la base de la mayoría de estándares a tratar en este apartado.

24.7.3. Estándares para representación y obtención de información geográfica

Entre los estándares más importantes encontramos aquellos que especifican la forma de recoger la información geográfica, así como aquellos que definen el modo en que esta se transmite.

Los siguientes estándares OGC forman parte de este grupo.

Simple Features for SQL (SFS)

Sabemos del capítulo 12 que el lenguaje SQL en su forma básica no sirve para recoger las geometrías que forman la parte espacial de una entidad, sino únicamente los datos no espaciales de esta. Sin embargo, versiones posteriores de SQL permiten la definición de tipos personalizados, y esto puede emplearse para poder incorporar estos elementos espaciales dentro del lenguaje.

El problema que nos encontramos en este caso es que, debido a que la propia flexibilidad de este mecanismo, los tipos pueden implementarse de diversas formas, lo cual no favorece la interoperabilidad. Si una consulta se establece sobre unos tipos definidos de forma distinta a como lo están en la base de datos que recibe la consulta, esa consulta no podrá procesarse correctamente. Es necesario definir una forma estandarizada de definir esos tipos, y una pauta a seguir para su implementación.

OGC define la especificación *Simple Features for SQL* (SFS) con objeto de hacer frente al problema anterior. SFS define por un lado unos tipos estandarizados para geometrías, los cuales se basan en otra especificación OGC denominada *OpenGIS Geometry Model*, que establece una forma de definir geometrías. Por otra parte, se definen una serie de operaciones SQL que operan sobre esos tipos.

Todas las geometrías que pueden definirse según este esquema son geometrías en un espacio bidimensional, y cada objeto geométrico está asociado a un sistema de referencia en el cual se define.

[3] Lenguaje de Marcado Extensible

Existe un objeto fundamental denominado *Geometry* del que heredan los restantes en una jerarquía bien definida. Los métodos de este objeto son de tres tipos:

- Métodos básicos. Proveen información sobre el objeto (dimensión, tipo de geometría, sistema de referencia, etc.)
- Métodos para comprobar relaciones espaciales entre objetos geométricos (cruza a, contiene a, se intersecta con, etc.)
- Métodos que efectúan algún tipo de análisis (unión de geometrías, distancia entre geometrías, área de influencia de una geometría, etc.)

Cada uno de los objetos derivados de la clase raíz *Geometry* tiene además a su vez sus propios métodos específicos, siempre dentro de alguno de los grupos anteriores.

Con estos objetos y sus métodos se da respuesta a todas las necesidades que aparecen en la realización de consultas sobre bases de datos espaciales. La especificación SFS permite así dotar de potencia al lenguaje de consulta SQL y hacerlo de forma estandarizada para ampliar la interoperabilidad en las operaciones de consulta.

Geography Markup Language (GML)

El *Geography Markup Language* (GML) es un lenguaje basado en XML, diseñado para el almacenamiento de información geográfica. Utilizando este lenguaje, resulta posible el intercambio de información geográfica de forma interoperable.

GML puede utilizarse para transmitir información a través de una red, como parte de un servicio. Este es el caso del servicio WFS que veremos más adelante, que devuelve información geográfica codificada según este lenguaje. No obstante, puede emplearse igualmente para almacenar la información con la que trabajamos de un SIG, del mismo modo que utilizamos cualquiera de los formatos de archivo que vimos en el capítulo 6. Es decir, sin que tengan que mediar servicios en ningún momento.

Algunos SIG permiten este uso, y soportan GML como un formato más de archivo. No obstante, no es una práctica común, ya que, pese a las ventajas de ser un estándar aceptado, GML es un formato de fichero de tipo texto (está basado en XML) y produce archivos de gran tamaño. Para este uso, es más habitual recurrir a algún otro formato.

GML es un lenguaje extremadamente genérico, que permite recoger tanto datos ráster como vectoriales y hacerlo con mucha flexibilidad. Permite, por ejemplo, recoger datos vectoriales sin que exista una geometría asociada, es decir, simplemente almacenando unos atributos como si se tratara de una base de datos no espacial. Esta gran flexibilidad, que es uno de los puntos fuertes de GML, es también uno de sus inconvenientes, ya que la especificación es muy compleja y difícil de implementar en su totalidad.

La versión más reciente de GML es GML3.

Existe un dialecto conocido como *Simple Features Protocol* que trata de solucionar el problema de la excesiva complejidad de GML3.

Web Feature Service (WFS)

El servicio *Web Feature Service* WFS está relacionado con los datos de tipo vectorial, y a través de él se sirven directamente las entidades de un dato vectorial con sus geometrías y datos alfanuméricos asociados. Desde este punto de vista, acceder a un servicio WFS es similar a acceder a una capa vectorial cualquiera o a una base de datos, ya que el SIG puede recuperar la información correspondiente (tanto la componente geográfica como la temática de cada entidad) y operar con ella.

En particular, las operaciones que permite un servicio WFS son:

- Crear una nueva entidad.
- Borrar una entidad.
- Actualizar una entidad.
- Obtener o consultar el conjunto de entidades en base a condiciones espaciales y no espaciales.

Distinguimos dos tipos de servicios WFS:

- Un servicio WFS básico, que solo permite consultar los datos, pero no modificarlos.
- Un servicio WFS transaccional (WFS-T), que permite realizar modificaciones en las entidades.

La versión más actual de la especificación WFS es la 1.1. No obstante, la versión 1.0 es la implementada mayoritariamente en los servidores actuales. WFS 1.1 utiliza GML3 como lenguaje para la codificación de la información a servir, mientras que WFS 1.0 usa GML2.

Filter Encoding

Cuando un cliente efectúa una petición a un servidor WFS, no es necesario que obtenga de este todas las entidades de una capa. Incluso para una zona geográfica dada, el usuario puede querer obtener a través del cliente solo aquellas entidades que cumplan un criterio dado.

Ya conocemos elementos que permiten realizar ese tipo de consultas para trabajar con un subgrupo de las entidades de una capa. En el capítulo 12 vimos el lenguaje SQL, mediante el cual podían definirse consultas de esta clase.

El estándar Filter Encoding define un formato basado en XML para el almacenamiento de expresiones de filtrado según otro estándar OGC conocido como *OGC Common Catalog Query Language*. La expresión del filtro expresada según la especificación Filter Encoding puede ser validada y procesada por herramientas adicionales para convertirla en las expresiones correspondientes en otro lenguaje para consulta de bases de datos espaciales. Por ejemplo, en una clausula `WHERE` de SQL que emplear en una sentencia `SELECT`.

Las expresiones que pueden recogerse empleando *Feature Encoding* pueden ser consultas con componente espacial o hacer referencia a la parte temática de la información geográfica. Es decir, que permiten recoger toda la variabilidad de las consultas espaciales que vimos en el capítulo 12

Además de emplear estas expresiones para consultar servicios WFS, pueden utilizarse igualmente para otros como los servicios de Nomenclátor (Gazetteer) que veremos más adelante, y en general en todos aquellos en los que tenga sentido especificar algún tipo de restricción a la hora de realizar una petición al servidor.

Web Coverage Service (WCS)

Si el estándar WFS permite obtener de un servidor datos vectoriales en forma de entidades, el estándar *Web Coverage Service* hace lo propio con datos ráster. Este servicio está pensado para tratar con *coberturas*, es decir, representaciones de un fenómeno que varía en el espacio. Como ya vimos en su momento, las coberturas se corresponden con el modelos de campos.

Representar una cobertura puede hacerse de muchas formas distintas: capas ráster, Redes de Triángulos Irregulares (TIN) o funciones matemáticas. No obstante, por el momento el estándar WCS solo está preparado para el trabajo con mallas ráster regulares.

EL servicio WCS ofrece los datos de la capa ráster como tales, con su semántica original. Es decir, que un servicio WCS puede servir un MDE y el cliente obtiene directamente los valores de elevación en sus unidades correspondientes.

24.7.4. Estándares para mapas y visualización

De entre todos los estándares que vamos a ver en esta sección, los más importantes por ser los

más extendidos son los que sirven mapas. Entendemos por *mapa* en este contexto a una representación gráfica de una determinada información geográfica, elaborada a partir de una o más capas.

Gran parte de los sitios Web que ofrecen información geográfica lo hacen en forma de mapas, es decir, que permiten simplemente «ver» los datos geográficos, y los estándares de esta sección son los encargados de definir ese tipo de servicios.

El estándar WMS, el principal en esta categoría, está ampliamente probado e implementado en gran cantidad de software, y es el soporte fundamental para cientos de aplicaciones basadas en mapas, lo que ratifica su utilidad y validez.

Web Map Service (WMS)

El estándar *Web Map Service* (WMS) define los elementos necesarios para un servicio de mapas.

Un servicio WMS devuelve una imagen con información geográfica, pero esta solo contiene la propia información visual para que el cliente pueda mostrarla. Es decir, si se pide a este servicio un mapa creado a partir de un MDE, la información de los píxeles no contiene la elevación de la coordenada correspondiente, sino el color asociado en función de un determinado criterio. La imagen puede contener otros elementos visuales tales como etiquetas o símbolos, en función de cómo se haga la representación en el servidor. Una vez que el cliente recibe la imagen, no puede actuar sobre esta para cambiar la forma de representación de una capa, sino simplemente representarla como es.

En un servicio WMS, cuando el cliente pide un mapa al servidor, puede controlar en cierto modo la forma en que este va a representarlo (colores, estilos, etc.). El servidor ofrece una serie de opciones predeterminadas, de las cuales el cliente solo conoce su nombre, y puede elegir una de ellas. No obstante, no puede saber exactamente qué caracteriza a cada uno de esos perfiles predeterminados de representación ni tampoco puede definir los suyos propios.

Para solucionar esto y ampliar las capacidades del servicio WMS, aparece otro nuevo estándar: SLD.

Styled Layer Description (SLD)

El estándar OGC *Styled Layer Description* (SLD) define una forma de almacenar los parámetros de representación empleados para crear un mapa a partir de los datos geográficos. Este estándar permite extender las capacidades de WMS, ofreciendo al cliente la posibilidad de definir sus propias configuraciones.

SLD es un estándar complejo que permite cubrir situaciones variadas y no solo las más sencillas y habituales. Permite, por ejemplo, el ajuste de elementos tales como etiquetas o simbologías personalizadas para elementos puntuales (por ejemplo, representar cada punto de una capa de localizaciones de estaciones de autobús con un pequeño dibujo de un autobús), Para esto último se apoya en otros estándares tales como SVG, diseñado para la representación de gráficos vectoriales.

Las simbologías recogidas en un documento SLD pueden emplearse para la representación tanto de capas ráster como vectoriales.

A la hora de definir una simbología para una capa, es necesario conocer cierta información acerca de esta. Para definir una simbología sencilla en la que todos los elementos de una capa van a ser representados de igual forma (por ejemplo, todos las líneas de una capa de ríos con un grosor dado y en color azul), esta información no es imprescindible, pero en caso de que se quiera variar ese color o ese grosor en función de un atributo, será necesario conocer qué atributos tiene la capa y de qué tipo son.

Para hacer esto, pueden emplearse las operaciones de los servicios de donde se toman los datos a representar. La operación *DescribeLayers* del servicio WFS permite conocer los tipos de entidades de una capa representada. La información sobre los atributos puede obtenerse con la operación *DescribeFeatureTypes*.

Web Mapping Context (WMC)

El estándar *Web Mapping Context* (WMC) define un formato estandarizado para almacenar un *contexto*. Un *contexto* recoge la información necesaria para reproducir las condiciones de una determinada sesión de uso de un cliente, de tal forma que ese cliente pueda restablecerlas posteriormente. El contexto se almacena en un archivo XML.

En el contexto se almacena información sobre las capas que forman el mapa representado por el cliente y los servidores de los que estas se obtienen, la región cubierta por el mapa, así como información adicional para anotar este mapa.

Los usos que se le pueden dar a un contexto son variados, entre ellos los siguientes:

- Mediante un contexto se puede definir una configuración particular de inicio para distintos tipos de usuario del cliente.
- Un contexto puede emplearse para almacenar el estado del cliente a medida que el usuario navega y modifica elementos, pudiendo retornar a una configuración establecida anteriormente.
- El contexto puede almacenarse y después transferirse a otro cliente distinto en el que comenzar en una misma configuración.

Los contextos pueden a su vez catalogarse y descubrirse, ofreciendo así un nivel de granularidad más amplio que las capas individuales. Pueden crearse diferentes contextos predefinidos y después hacer estos accesibles para facilitar el establecimiento de una determinada configuración en un cliente.

24.7.5. Estándares para metadatos, catálogos y consulta de datos

Los metadatos y las operaciones sobre ellos tienen sus propios estándares bien definidos.

Por una parte, existen estándares dedicados a los metadatos en sí y a la forma de almacenarlos. Estos pueden especificar parámetros relativos a los metadatos tales como los siguientes:

- Contenido de los metadatos, definiendo qué campos son obligatorios y cuáles opcionales.
- Formato de almacenamiento. En general, una descripción del formato a emplear.
- Prácticas adecuadas de creación y actualización. Se definen las pautas correctas que han de seguirse a lo largo del ciclo de vida de los datos.
- Reglas de conformidad. Reglas que permiten comprobar si un determinado metadato se encuentra conforme con el estándar.

Por otro lado, un conjunto de metadatos conforma la base para las consultas sobre un catálogo, el cual describe a su vez un conjunto de datos. Como ya vimos, el catálogo constituye una forma más sencilla y eficaz de consultar los datos, agilizando las operaciones y permitiendo el descubrimiento de datos de forma óptima, por lo que la consulta de estos metadatos también debe estar estandarizada, y debe definirse cómo los clientes deben obtener la información de los metadatos para posteriormente, a partir de dicha información, realizar el acceso a los datos correspondientes que resulten de interés.

ISO 19115 e ISO 19119

ISO 19115 e ISO 19119 son los estándares ISO para metadatos asociados a información geográfica. Definen más de 400 elementos, de los cuales

los siguientes forman parte de su núcleo fundamental.

- Título
- Fecha de referencia de los datos
- Idioma
- Categoría en que encuadrar la temática de los datos
- Resumen
- Punto de contacto para los metadatos
- Fecha de los metadatos
- Organismo responsable de los datos
- Localización
- Juego de caracteres de los datos
- Resolución espacial
- Formato de distribución
- Tipo de representación espacial
- Sistema de referencia
- Recurso en línea
- Identificador del fichero de metadatos
- Nombre del estándar de metadatos
- Versión del estándar de metadatos
- Idioma de los metadatos
- Juego de caracteres de los metadatos

En España, existe el Núcleo Español de Metadatos (NEM), un subconjunto de la ISO 19115 definido por un subgrupo de trabajo de la Infraestructura de Datos Espaciales de España (IDEE).

Nomenclátor (Gazetteer)

Un *nomenclátor* o *gazeteer* permite la localización de fenómenos geográficos a partir de un determinado nombre. El catálogo sobre el que se basa es una colección de estos fenómenos, cada uno de ellos asociados a un identificador geográfico. Dicho identificador es una referencia espacial en forma de etiqueta o código que identifica un lugar en el mundo real. Ejemplos de tales identificadores son los nombres de ciudades o pueblos (Burgos, Plasencia), los códigos postales (10600), los accidentes geográficos (Puerto de Navacerrada, Pico de la Miel) o las direcciones (Carretera N-V p.k.35, Calle Mayor 32), entre otros. Así, el servicio de nomenclátor permite establecer un sistema de referencia basado en identificadores geográficos.

El servicio recibe como entrada un nombre y utiliza este para localizar los fenómenos geográficos que cumplen un criterio. Este criterio puede ser variable, pudiendo exigir que el nombre coincida plenamente, que comience por él, o que lo contenga, entre otras opciones. Es habitual además que el catálogo contenga una tipología de los fenómenos recogidos (población, río, puerto, lago, etc.), de forma que esta también puede utilizarse para establecer el criterio de consulta (por ejemplo, para localizar todos los ríos que comiencen con la letra «b»).

En el terreno de los nomenclátor encontramos la Norma ISO 19115:2003 (*Geographic Information – Spatial referencing by Geographic Identifiers*), y los *OGC Catalog Services*, que permiten estandarizar procesos de consulta como los mencionados.

24.7.6. Estándares para procesamiento

Además de servir datos, pueden servirse procesos sobre esos datos, de tal forma que existan procesos remotos a los que los clientes pueden acceder. También debe estandarizarse la forma de acceso a estos servicios y cómo los clientes efectuarán las peticiones de procesos y la transmisión o definición de los datos que han de tomarse para esos procesos.

24.7.7. Web Processing Service (WPS)

El estándar *Web Processing Service* (WPS) de OGC está enfocado a definir el marco en el que se ha de producir el servicio de procesos remotos. WPS define una interfaz estándar que facilita la

publicación de procesos y su uso posterior por parte de clientes. Se entiende por proceso en este contexto a todo aquel algoritmo, cálculo o modelo que opere sobre datos georreferenciados.

Los procesos que pueden definirse son sumamente flexibles, pudiendo tener un número cualquiera de entradas y salidas, y operar con distintos tipos de datos. Es decir, que ofrece un marco para definir cualquier tipo de proceso de análisis geográfico, tanto si este utiliza datos ráster como si utiliza datos vectoriales. Todos los procesos que hemos visto en la parte correspondiente al análisis espacial pueden ofrecerse como servicios remotos a través de WPS.

Los datos empleados para alimentar los procesos pueden encontrarse en el propio servidor o ser transmitidos a través de la red al igual que la propia petición de proceso por parte del cliente. Asimismo, puede relacionarse este estándar con otros que ya hemos visto, como por ejemplo WFS. Los datos necesarios para ejecutar un proceso que requiera una capa vectorial pueden obtenerse llamando a un servicio WFS, en cuyo caso debe indicarse en los parámetros del proceso los propios parámetros que corresponden a la petición a ese servicio WFS.

24.7.8. Relación entre estándares

Los estándares que hemos visto a lo largo de estás páginas guardan una lógica relación entre ellos. Dentro de un mismo ámbito, los estándares pueden guardar relación con otros similares aun habiendo sido desarrollados por entidades distintas. El objetivo de armonización tecnológica que pretenden los estándares resulta más difícil de lograr si el número de estándares para una misma tecnología es elevado, ya que los fabricantes necesitan dedicar más tiempo y recursos a implementar todos ellos, y lo normal es que opten por implementar solo algunos.

Por este motivo, las organizaciones que promueven estándares trabajan conjuntamente y suelen producir estándares muy similares. En algunos casos, como ya hemos mencionado, algunos estándares OGC son también estándares ISO, existiendo no una similitud sino una absoluta coincidencia.

Más importante es la relación que guardan entre sí estándares dedicados a áreas distintas. Las tecnologías para la gestión y transmisión de datos incluyen diversos elementos que forman un todo interrelacionado. Los estándares correspondientes a esos elementos y a cada proceso particular que se desarrolla deben formar también un todo conectado y poder a su vez «entenderse» con otros estándares relacionados.

Un caso particular de esto es, por ejemplo, el de los estándares WMS, SLD y WFS. El servicio WMS ofrece un mapa, que no es sino una representación de unos datos según unos criterios dados. Esos datos pueden obtenerse de un servicio WFS y los criterios para representarlos pueden expresarse utilizando el estándar SLD. La ventaja de los estándares abiertos, máxime si estos han sido además creados por una misma organización, es la capacidad de interoperar entre ellos, de forma que WMS puede tomar datos de servicios WFS o WCS, o una consulta conforme a Filter Encoding puede aplicarse para consultar un servicio WFS y también un servicio de nomenclátor.

Otro ejemplo en esta línea es el que hemos descrito para un servicio WPS que toma datos de un servicio WFS para operar con ellos.

En su conjunto, deben verse todos los estándares como una gran familia de elementos que armoniza el trabajo con la información geográfica.

24.8. Resumen

Hemos visto en este capítulo las ideas fundamentales del binomio cliente-servidor, tanto en su definición más general referente a servicios Web de cualquier tipo, como en aquellos específicos del ámbito SIG. En base a esto, existen distintas formas de llevar a la red tanto los propios datos geográficos como las funcionalidades principales de los SIG de escritorio, y que pueden variar en cuanto a su complejidad, desde simples mapas estáticos hasta aplicaciones Web complejas.

Pese a las elevadas posibilidades que existen hoy en día en cuanto a tecnologías Web, es importante conocer también las limitaciones del entorno de trabajo, las cuales derivan tanto de la propia red como de otros aspectos, por ejemplo el hecho de que la aplicación Web se ejecute dentro de un navegador. Estas limitaciones llevan al desarrollo de técnicas particulares para optimizar el funcionamiento de las aplicaciones SIG Web, entre las que se han de destacar el *tiling* y el *cacheo*.

Asimismo, conocemos ya las funcionalidades principales que debe presentar un servidor para responder a las peticiones de un cliente SIG, que son principalmente servir representaciones de los datos geográficos, servir los datos en sí o consultas sobre estos, o bien servir procesos de análisis basados en dichos datos.

Con el fin de garantizar la fluidez en las operaciones dentro del sistema cliente-servidor, aparecen los estándares, los cuales permiten que los elementos de este sistema sean interoperables.

SIG móvil

25.1. Introducción

En el veloz avance que el *hardware* sufre constantemente, uno de los cambios más radicales de los últimos tiempos es la cada vez mayor potencia y disponibilidad de elementos portátiles. Esto ha propiciado la aparición del denominado *SIG móvil*, así como una serie de tecnologías y herramientas relacionadas que van dando forma a un sector muy distinto de lo que el SIG clásico representa, pero con una innegable vinculación con este.

La implicación que estas nuevas tecnologías han tenido en el ámbito del SIG va más allá de expandir sus posibilidades. Como vimos en el capítulo dedicado a la historia de los SIG, los primeros programa SIG se ejecutaban sobre grandes máquinas cuya adquisición estaba muy lejos del alcance del público especializado, como sucedía con toda la tecnología informática de aquel entonces. El salto a los ordenadores personales fue decisivo para iniciar una popularización de los SIG y contribuir a que se convirtieran en herramientas imprescindibles en una buena parte de sus ámbitos de aplicación, entrando con fuerza en muchos sectores.

Con la aparición de los dispositivos móviles y el crecimiento del mercado en torno a ellos, los SIG han dado un nuevo salto cualitativo. No solo han alcanzado un nuevo tipo de dispositivos con capacidades valiosas relacionadas con la información geográfica (destacando entre ellas la de conocer la posición del dispositivo), sino también a un nuevo público y a nuevos grupos de interés. Si con el salto a los ordenadores personales los SIG se hicieron más asequibles en términos económicos y de especialización informática, con la entrada de los dispositivos móviles se han hecho asequibles en lo que a conocimientos específicos del ámbito geográfico y cartográfico respecta. La información geográfica se abre paso en un mercado no especializado y, no solo su uso, sino también su creación, pasan ambos a ser actividades no exclusivas de los profesionales de este campo. Es un paso más allá en la labor que desde sus orígenes los SIG vienen realizando. Esto es, facilitar el uso de información geográfica y dar presencia a esta en todos los terrenos.

25.2. Qué es el SIG móvil

Comencemos viendo qué entendemos por SIG móvil y de qué formas puede presentarse. Para ello, analicemos como hemos hecho en otros apartados una situación habitual en el entorno de trabajo de un SIG, en particular la relativa a la creación de datos geográficos. Según vimos en el capítulo 6, una de las formas de obtener datos con los que trabajar en un SIG es la toma de estos directamente en campo y su posterior incorporación dentro del SIG. Esto requería habitualmente la digitalización de los datos tomados, ya que di-

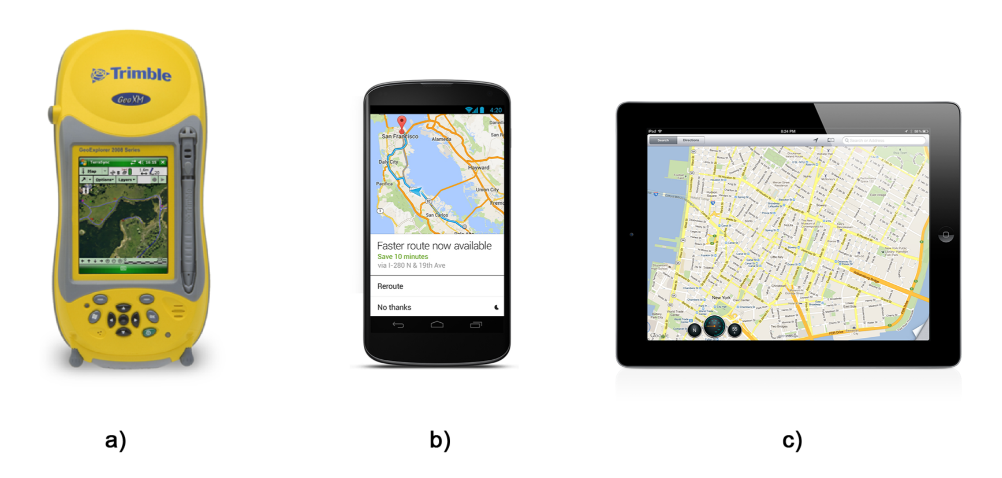


Figura 25.1: Distintos tipos de dispositivos móviles

cha toma se desarrollaba en la mayor parte de los casos mediante medios analógicos.

Una importante mejora en este proceso se daría si la recogida de datos se efectuara empleando medios digitales, ya que el resultado sería mucho más cercano a lo que posteriormente va a necesitarse para el trabajo en gabinete. Además de esta ventaja inmediata y del ahorro de tiempo que traería consigo, existen otras ventajas directas que los medios digitales aportarían. Por ejemplo, si en lugar de un dispositivo de toma de datos contamos con dos de ellos, es relativamente sencillo (especialmente si los dispositivos pueden comunicarse entre sí) replicar los datos tomados, teniéndose así una copia de seguridad que evitará en gran medida la pérdida de estos. El software instalado en el dispositivo puede a su vez contar con elementos que efectúen algún tipo de control de calidad, asegurándose de que no se introducen por descuido valores erróneos o de que no se deja sin rellenar ningún campo de un estadillo.

En este supuesto, y pesar de que esa recogida de datos se realiza con el fin de tratar estos posteriormente con ayuda de un SIG, el SIG como tal no participa en el proceso de toma de datos. Los dispositivos que se lleven al campo, sean del tipo que sean, simplemente recogeran una serie de valores introducidos por el operario correspondiente, del mismo modo que este los anotaría en una libreta o estadillo. Sin embargo, nada impide que podamos llevar al campo una aplicación SIG (por ejemplo, un SIG de escritorio), y que sea en ese SIG, adaptado a esa particular tarea o no, donde tomemos los datos y en el que nos apoyemos para realizar todas las tareas adicionales que el trabajo de campo requiere.

Llevar el SIG al campo y sacarlo de su lugar fijo en el gabinete es lo que da lugar al denominado *SIG móvil*, cuyas particularidades, como veremos, van más allá del mero hecho de una localización de trabajo distinta. Cuando hablamos de SIG móvil, no nos referimos únicamente a un SIG habitual ejecutándose en una plataforma móvil, sino también a una filosofía distinta a la que existe en el uso de otros elementos tecnológicos del ámbito SIG que ya conocemos, en una localización fija.

Hoy en día son muchos los dispositivos que podemos emplear para disponer de una herramienta móvil con capacidad de proceso. Sin ir más lejos, un ordenador portátil con una conexión inalámbrica a Internet (empleando telefonía

móvil de tercera generación) nos serviría para replicar en campo el entorno de trabajo de un SIG de escritorio, poniendo a nuestra disposición todas las capacidades de este. Esta solución, no obstante, es poco práctica, ya que, si bien es cierto que podemos «mover» un ordenador portátil con relativa facilidad, no es un dispositivo pensado para moverse mientras se encuentra en funcionamiento, con lo que más bien tendríamos un *SIG portátil* en lugar de un *SIG móvil.* Hay otros tipos de *hardware* mucho más adecuados para este cometido, como pueden ser los siguientes (Figura 25.1):

- Unidades GPS (a).
- Teléfonos móviles (b).
- Tabletas (c).

Las características de estos dispositivos son distintas a las de un ordenador de sobremesa en el que utilizamos el *software* SIG que hemos visto hasta ahora, haciendo que deba desarrollase software específico y que deban tenerse en cuenta algunas consideraciones adicionales. A su vez, cada uno de estos dispositivos tiene unas capacidades propias que lo hacen más interesante para unas u otras tareas dentro del trabajo en campo.

Así, las tabletas o los teléfonos moviles de tipo *smartphone* pueden considerarse como versiones reducidas de un ordenador de sobremesa o un ordenador portátil, y aunque en términos de capacidad de almacenamiento y velocidad de proceso están por debajo de estos, son dispositivos de gran potencia que en muchos casos pueden ejecutar aplicaciones complejas o que requieran la realización de procesos intensos.

Por su parte, los teléfonos móviles son los dispositivos versátiles por excelencia y su penetración es muy superior a la de cualquier otro.

Las unidades GPS más básicas se limitan a mostrar la localización, disponiendo de funcionalidades reducidas. Las más completas, no obstante, incorporan capacidades más cercanas a las de una tableta o *smartphone*, con posibilidad de ejecutar aplicaciones complejas tales como un SIG adaptado. El interés de la tecnología GPS está, sin embargo, en considerarla como una tecnología adicional que enriquece a algunos de los dispositivos anteriores. Así, tanto teléfonos móviles como las tabletas (o incluso otros dispositivos como cámaras fotográficas) pueden incorporar receptores GPS y disponen por tanto de información acerca de su posición. Esta combinación es la que da como resultado los dispositivos más potentes para el SIG móvil, ofreciendo todas las funcionalidades que iremos viendo a lo largo de este apartado.

La conexión remota a Internet abre la puerta a muchas de las capacidades más potentes y novedosas del SIG actual, como pueden ser la consulta o incluso la edición de cartografía, según vimos en el capítulo 24.

Para dar una definición más formal de lo que entendemos por SIG móvil, podemos decir que es una tecnología que integra una o más de las siguientes [317]:

- Dispositivos móviles.
- Sistemas de posicionamiento global (GPS).
- Acceso inalámbrico a Internet.

Por su parte, [318] distingue tres elementos principales que dan forma al contexto de las aplicaciones SIG móviles: SIG, Internet, y dispositivos móviles y Nuevas Tecnologías de la Información y la Comunicación (NTIC).

Los denominados *Servicios Basados en Localización* (LBS[1]) aglutinan todas estas tecnologías. En general, se suelen recoger bajo esta denominación los servicios que toman en consideración la posición del usuario, y en los que se produce la participación de un tercero, el encargado de

[1] *Location–Based Services*

proveer el servicio como parte fundamental de un negocio. Dichos servicios pueden ir desde la localización del comercio más próximo hasta el envío de avisos cuando se encuentre cerca de otro usuario conocido.

Podemos, con lo anterior, tener así una primera clasificación general de las áreas de aplicación del SIG móvil en los dos siguientes grupos:

- SIG «en campo». Se centra más en los trabajos propios del SIG y en la recolección y edición de datos.
- Servicios Basados en Localización. Servicios ofrecidos por terceros en función de la posición del dispositivo y del usuario.

En los LBS, la persona con el dispositivo es consumidor del servicio, mientras que en el SIG en campo su papel es principalmente como operario del SIG, y por tanto es esa persona la que provee un servicio o realiza una tarea apoyado en él. Se tiende a concebir el LBS como un servicio no especializado cuyo consumidor no ha de estar necesariamente formado en las tecnologías SIG, mientras que en el caso del SIG en campo sí que debe tener unos conocimientos mínimos, ya que su labor se desempeña sobre una aplicación SIG como tal. De los elementos que hemos comentado como integrantes del SIG móvil, el LBS da mayor importancia al acceso a Internet y a la posición del dispositivo, dejando algo más de lado las capacidades clásicas del SIG. El SIG en campo, por su parte, hace énfasis en esas capacidades, complementándolas con la movilidad del dispositivo y su capacidad para calcular su posición.

En nuestro supuesto con el que comenzábamos esta sección, la toma de datos para incorporar estos posteriormente en un SIG, nos encontraríamos en un claro caso de SIG en campo. Este tipo de enfoques surgieron antes que los LBS, ya que las tecnologías necesarias para estos últimos aparecieron con posterioridad. El SIG en campo no requiere obligatoriamente una conexión inalámbrica, tecnología de muy reciente aparición y, sobre todo, de muy reciente implantación y desarrollo. La llegada de esta tecnología, sin embargo, añadió nuevos elementos al SIG móvil, y a día de hoy es la cabeza visible de este ámbito, especialmente por la gran expansión que ha supuesto para las tecnologías SIG. Como mencionábamos en la introducción del capítulo, la popularización del SIG y sus elementos es el verdadero aspecto destacable del SIG móvil.

Pese a lo anterior, la frontera entre estos dos grupos es difusa en cierto modo, ya que puede realizarse trabajo de campo aprovechando servicios de terceros a través de Internet, y el usuario que aprovecha estos servicios (que pueden a su vez ser muy especializados) puede tener amplios conocimientos de SIG y realizar un trabajo altamente técnico. En relación con cuanto hemos visto en otras partes del libro, el SIG en campo está, a primera vista, más vinculado con todo ello, ya que el perfil de su usuario es más similar al del clásico usuario de SIG. La importancia que los LBS están teniendo es, aun así, mucho mayor, ya que alcanzan a todo el ámbito del SIG y también a grupos de usuarios muy alejados de ese perfil tradicional.

Independientemente de la naturaleza de la actividad realizada con un SIG móvil, está claro que este tiene unas particularidades que lo diferencian del SIG como hasta ahora lo hemos conocido, y que son las que, en gran medida, le confieren su potencia específica como herramienta para trabajo sobre el terreno.

25.3. Particularidades del SIG móvil

Los siguientes son algunos de los principales aspectos que caracterizan al SIG móvil y lo

diferencian del SIG clásico sobre una plataforma estática [319]:

- **Variedad de plataformas**. Mientras que en en caso de un SIG que se ejecuta en un ordenador de sobremesa las diferencias de plataforma son prácticamente inexistente (con, tal vez, la única salvedad del sistema operativo), en el caso del SIG móvil la situación es muy diferente. Existen plataformas muy diversas y dispositivos con características completamente distintas (por ejemplo, un teléfono móvil es, en ciertos aspectos, radicalmente distinto a una tableta). Garantizar que todos estos dispositivos van a poder funcionar con una aplicación requiere un esfuerzo extra a la hora de desarrollar esta.
- **El usuario es parte de la información**. El SIG nos permite analizar información muy variada, pero los análisis que realizamos se basan en unos datos concretos, ya sean estos locales o remotos. La posición de la maquina donde se ejecuta el SIG no es relevante ni se tiene en cuenta, y ni siquiera existe la posibilidad de conocer y utilizar esta. En el SIG móvil, por el contrario, la posición del dispositivo es conocida (si este integra algún tipo de mecanismo para calcular está, de entre los que veremos más adelante en esta misma sección). Esa posición no solo puede ser empleada como otro dato más, sino que, en muchos casos, es el dato más importante y el que permite ofrecer servicios personalizados en función de dicha posición. Indirectamente, el usuario se convierte también en parte de la información, ya que es *su* posición la que ahora forma parte de esta.
- **Acceso variable**. La calidad del acceso a Internet va a fluctuar notablemente para un mismo conjunto de dispositivo, aplicación y usuario, ya que se trata de un servicio muy variable en función de la localización.
- **Limitaciones de los dispositivos**. Comparados con un ordenador de sobremesa, que representa el dispositivo estándar en el que un SIG se ejecuta tradicionalmente, los dispositivos móviles presentan importante limitaciones. Las más destacable de ellas es su propio tamaño, ya que las pantallas son pequeñas y obligan a un uso distinto de su espacio para poder mostrar en ellas todos los elementos que garantizan una correcta usabilidad de las aplicaciones. Otras limitaciones son las ya mencionadas de almacenamiento y proceso. Y, por último, deben considerarse también las limitaciones en los dispositivos de entrada, muy distintos de los habituales teclado y ratón, y sin apenas posibilidad de contar con otros periféricos más específicos.
- **Escalabilidad de los datos**. Por las propias características tanto de los dispositivos como de sus conexiones, es necesario poner atención en la escalabilidad de los datos para que las aplicaciones funcionen en circunstancias variadas, modificando el detalle en función de la situación.

25.4. Aplicaciones del SIG móvil

Para estudiar las posibilidades que el SIG móvil nos brinda, podemos analizar el papel que la información geográfica juega en el trabajo de campo. Entendemos aquí trabajo de campo como cualquier actividad desarrollada al aire libre en la que pueda aplicarse de algún modo un SIG móvil.

Por una parte, la información geográfica es una herramienta en la que nos apoyamos para desarrollar la actividad en cuestión. Es decir, *usamos* la información geográfica de forma directa. Así sucede, por ejemplo, cuando debemos tomar datos en una localización concreta como por ejemplo una parcela de inventario en un inventario forestal o un punto de alcantarillado para realizar un control del estado de una red de saneamiento. También hacemos un uso similar cuando buscamos el restaurante más próximo o quere-

mos encontrar el camino más rápido para tomar una carretera desde nuestro emplazamiento actual.

Tradicionalmente, la información geográfica se ha llevado al campo en forma de mapas impresos. Consultando estos se encontraba el lugar seleccionado y la forma de desplazarse hasta él. Empleábamos mapas topográficos para encontrar esa parcela de inventario, callejeros para localizar la alcantarilla o un mapa de carreteras para saber cómo desplazarnos en coche. Con el SIG móvil, la información geográfica «viaja» al campo en formato digital, almacenada dentro del propio dispositivo o bien accediendo mediante este a información remota a través de Internet. Esto ofrece ventajas tales como una mayor comodidad o como la posibilidad de tener varios dispositivos que compartan la cartografía. Es decir, varios técnicos que trabajen en campo pueden «llevar» el mismo mapa sin necesidad de tener varias copias de este, sino tan solo varias «copias» del dispositivo, que es por otra parte el mismo que emplearán para la toma de datos o para cualquiera de las restantes tareas de su trabajo.

Por otra parte, la información geográfica en sí puede ser parte de los datos recogida en campo. Es decir, es objeto de interés directo del trabajo de campo, y no solo un medio para realizar este. En este caso, los dispositivos móviles van a permitir recoger con más precisión cualquier tipo de dato espacial sobre el terreno, al mismo tiempo que facilitan la creación de dicho dato espacial o la edición de uno ya existente en función de lo observado. Se unen en este punto la capacidad del dispositivo para conocer las coordenadas de su localización y las capacidades de las aplicaciones SIG para edición de datos, así como las propias ventajas de los datos digitales en lo que a su actualización respecta (véase 6.2).

Esta es una de las razones principales del auge actual de los proyectos colaborativos para la creación de cartografía (véase 6.7). Los complejos y caros equipos empleados en la cartografía clásica pueden sustituirse en muchos casos por dispositivos simples como un teléfono móvil o una unidad GPS de consumo, ambos sencillos de manejar para el usuario no especializado. Este puede así tomar información geográfica y aportarla a algún proyecto comunitario, o bien guardarla para su uso personal.

Con las ideas anteriores, podemos localizar las principales tareas que el SIG móvil va a desarrollar en los distintos ámbitos de aplicación y dividirlas en dos bloques: aquellas que permiten a los usuarios optimizar su movilidad durante el trabajo de campo, y aquellas que facilitan el desarrollo de la labor en cuestión una vez que se ha posicionado correctamente.

Con respecto a las relacionadas con la movilidad, no se ha de pensar que estas se limitan a la localización de un emplazamiento puntual como se ha mencionado anteriormente, en lo que sería un uso casi exclusivo del sistema de posicionamiento del dispositivo. También el análisis, parte importante de un SIG, puede servir para mejorar el desplazamiento que el trabajo en campo conlleva. El cálculo de rutas es el principal ejemplo en este sentido, tal y como se implementa en los navegadores GPS, aunque no el único. Elaborar un plan de ruta en tiempo real puede ser útil para muchos profesionales, que pueden hacer uso de algoritmos como el del «problema del viajante» si estos se encuentran implementados en su SIG móvil.

Dentro de las actividades que facilitan la labor en campo son de especial interés las relacionadas con la captura de información geográfica, que se simplifica notablemente como ya hemos dicho. Asimismo, también debemos citar cualquier tipo de servicio al que pueda accederse mediante la conexión inalámbrica del dispositivo y proporcione información complementaria o algún tipo de

apoyo a la persona que opera con este. Y por último, no ha de olvidarse el análisis SIG como una herramienta con gran utilidad, ya que permitirá realizar procesos adicionales que pueden añadir nuevas posibilidades, tales como, por ejemplo, la validación en tiempo real de los datos recogidos.

La siguiente lista resume algunas de las actividades principales que pueden llevarse a cabo con un SIG móvil. Algunas de ellas pueden desarrollarse sin necesidad de contar con todos los elementos posibles (dispositivo, conexión inalámbrica y sistema de posicionamiento), aunque buena parte requieren el concurso de todos ellos.

- **Navegación**. Cálculo de ruta óptima entre dos puntos, guiado en interiores (centros comerciales, museos, etc.), aparcamiento guiado, gestión de tráfico. Una de las actividades más populares y extendidas.
- **Inventario**. Recogida de datos de cualquier tipo sobre el terreno. Cubre desde datos de inventarios forestales a prospecciones arqueológicas, pasando por datos censales o infraestructuras urbanas, entre muchos otros.
- **Información**. Paginas amarillas espaciales o guias de viaje virtuales. En general, cualquier servicio de mapas o de puntos de interés con posición (monumentos, tiendas, aparcamientos...) accesible desde un dispositivo móvil.
- **Emergencia**. Localización de situaciones de emergencia, asistencia a vehículos, optimización de asistencias y tiempos de respuesta. El usuario, ante una emergencia, puede conocer su posición e informar de ella, o bien a través de la red puede conocerse esta y emplearse para dar una respuesta óptima y una ayuda lo más eficiente posible.
- **Publicidad**. Anuncios basados en localización, indicación de negocios cercanos, promociones para comercios próximos. Existen algunas limitaciones para evitar la publicidad no deseada, pero si el usuario da permiso, puede recibir información sobre posibilidades comerciales en su entorno.
- **Seguimiento**. Tanto de personas como de productos, a lo largo de rutas predefinidas o no. También puede servir para monitorizar una actividad en las distintas localizaciones por las que pase el usuario. Por ejemplo, una compañía telefónica puede estudiar los patrones de comportamiento en lo que al acceso a la red respecta, según el emplazamiento desde el que se accede.
- **Gestión**. Por ejemplo, de infraestructuras, de instalaciones, o de flotas. El dispositivo puede ir sobre el elemento a gestionar o bien emplearse para llegar hasta él y efectuar allí algún tipo de control.
- **Ocio**. Buscadores de amigos o juegos con componente espacial, entre otros.

25.5. Métodos de posicionamiento

Uno de los elementos clave del SIG móvil es la capacidad de conocer la posición del dispositivo en todo momento. Como ya hemos visto, esa posición se incorpora como un dato más, y es de particular relevancia para realizar otras operaciones habituales del SIG o para ofrecer servicios de diversos tipos. Si el dispositivo en cuestión es una unidad GPS u otra clase de dispositivo con receptor GPS incorporado, está claro que dispone de un sistema para obtener su posición. Sin embargo, existen otras formas de que el dispositivo conozca su posición, y pueden emplearse de igual modo para obtener resultados similares en cuanto a las prestaciones que van a permitir.

Los métodos mediante los cuales puede determinarse la posición de un dispositivo pueden clasificarse en tres tipos, a saber:

- Introducción manual de la posición.
- Métodos basados en red.

- Metodos basados en terminal.

La introducción manual es el método más obvio y simple que, no obstante, puede implicar también el uso de algún tipo de tecnología y requiere algunas matizaciones. Además de introducir directamente en el dispositivo las coordenadas actuales de este, es posible establecer una posición mediante la denominada *geocodificación inversa*. En el capítulo 6.5.3 veíamos que mediante la geocodificación asignábamos coordenadas a un determinado elemento, que podía ser un punto dado o cualquier otro elemento susceptible de ser georreferenciado. Aplicando este razonamiento de forma inversa, y si disponemos una base de datos con un conjunto de esos elementos y sus coordenadas asociadas, podemos obtener estas últimas haciendo búsquedas en esa base de datos con el nombre del elemento. Es decir, podemos decirle al dispositivo que la posición actual es *Badajoz* o *Estadio Vicente Calderón* y él se encargará de convertir esa información en una coordenada numérica similar a la que se obtendría si tuviera instalado un receptor GPS o alguna otra tecnología similar.

Algunos servicios de consulta de los que presentábamos en el capítulo 24 permiten este tipo de operaciones, y devuelven coordenadas asociadas a un determinado fenómeno geográfico. En particular, los denominadas servicios de *Nomenclator* ya citados en el capítulo anterior, son los encargados de ello.

Con respecto a los dos tipos restantes, ambos se apoyan en una red de estaciones cuyas posiciones son conocidas. Los basados en red obtienen su posición mediante cálculos realizados en función de una señal emitida por el dispositivo. El método más habitual de esta clase es el empleado por los teléfonos móviles para calcular su posición en función del repetidor más cercano de entre los que le ofrecen cobertura.

Por el contrario, en los métodos basados en terminal es el propio dispositivo el que recibe la señal que procede de las estaciones, y en función de estas calcula su posición. El sistema GPS es el ejemplo más popular de un método de esta última clase. Existen asimismo métodos combinados que emplean ambas técnicas para el cálculo posicional.

La figura 25.2 esquematiza lo anterior.

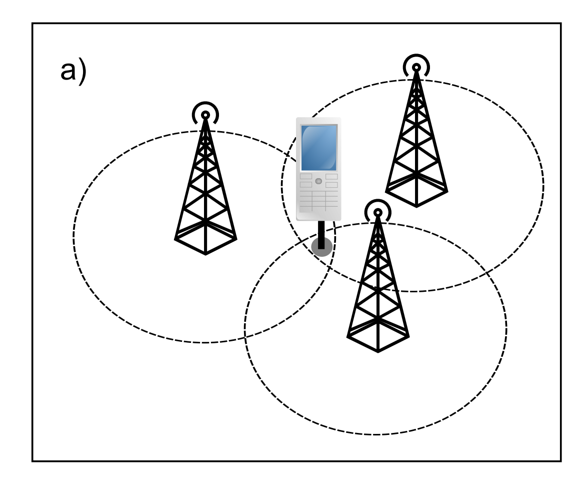
Con independencia del tipo de método, el proceso de cálculo de posición sigue un esquema como el siguiente:

- La posición de las estaciones es conocida.
- La información de la señal se transforma en una distancia (a excepción de si se aplica la técnica conocida como Ángulo de Llegada, que veremos seguidamente).
- La posición se calcula conociendo las distancias a un número dado de estaciones base.

Esto coincide con lo que ya vimos en el apartado 6.6 dedicado al sistema GPS.

Para convertir la información de la señal en una posición, encontramos diversas técnicas, a saber:

- Celda de Origen (Cell of Origin, COO). Se identifica la estación base más cercana y con ello se sabe que el dispositivo se encuentra en el perímetro de esta, dentro de su radio de alcance. La precisión depende de la densidad de la red. Para el caso de telefonía móvil, se sitúa entre los 200 metros y varios kilómetros, por lo que es baja para cierto tipo de servicios.
- Tiempo de Llegada (Time of Arrival, TOA). Se conoce la velocidad de transmisión de la señal y el tiempo entre el envio y la recepción de la señal, con lo que puede calcularse la distancia. Se tiene así la distancia respecto a una estación dada. Considerando la velocidad de transmisión de la señal, son necesarios relojes de alta precisión para lograr un calculo preciso.



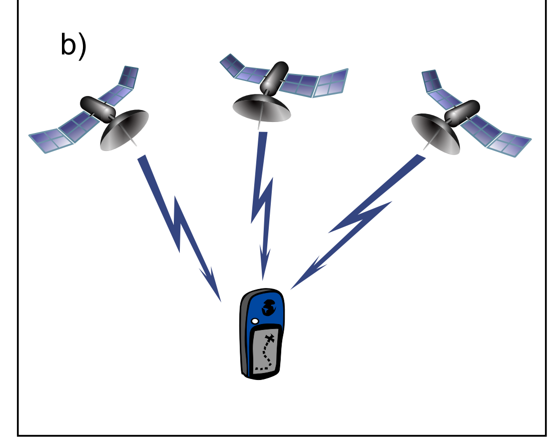


Figura 25.2: Metodos de posicionamiento basados en red (a) y en terminal (b)

- Diferencia de Tiempo de Llegada (Time Difference of Arrival, TDOA) o Diferencia de Tiempo Observada Mejorada (Enhanced Observed Time Difference, E-OTD). En ambas técnicas se mide igualmente el tiempo, pero el cálculo de la distancia se basa en la diferencia de las señales de tres estaciones, pudiéndose así triangular la posición. En el caso de TDOA el cálculo de la posición lo realiza el proveedor de la red, mientras que en el E-OTD es el dispositivo móvil quien lo hace.
- Ángulo de Llegada (Angle of arrival, AOA), Dirección de Llegada (DOA): Se usan antenas direccionables para detectar el ángulo de llegada.

Estas técnicas pueden emplearse simultáneamente, con objeto de proporcionar una localización más fiable o de adaptarse a las propias circunstancias de la red de estaciones en cada momento.

Es interesante mencionar que la precisión en los métodos basados en terminal es en general mayor que la de los métodos basados en red, siendo así más adecuados para servicios en los que la posición deba conocerse de forma más precisa [320]. Así, el GPS ofrece precisiones mucho mayores que las que se pueden obtener con la identificación de la celda más cercana en una red de telefonía móvil. El GPS es, sin embargo, una técnica pensada para emplearse en exteriores, y los servicios en interior no pueden hacer uso de este, además de requerir una precisión aún mayor. En este caso, métodos basados en redes locales inalámbricas (WLAN), Bluetooth o ultrasonidos son una opción válida, todos ellos también basados en terminal.

25.6. Redes inalámbricas

Uno de los elementos más importantes en el SIG móvil es la conexión inalámbrica, que nos permite el acceso a Internet y poder acceder a todos los tipos de servicios a través de esta. Sin conexión, disponemos de gran cantidad de funcionalidades, en especial aquellas fundamentales para lo que denominábamos SIG en campo. Podemos llevar el SIG móvil y tomar datos, realizar análisis geográficos sobre el terreno o navegar hasta una posición dada. Para ello solo necesitamos los datos que estén almacenados en el propio dispositivo, tal y como sucede en un navegador GPS que contiene su propia cartografía.

Sin conexión a Internet, sin embargo, no se dispone de capacidad para recibir servicios ni tampoco para acceder a datos remotos o realizar consultas sobre datos de terceros, limitando así de forma notable el alcance del SIG móvil.

Existen dos esquemas principales para clasificar las redes inalámbricas: según la topología de la red y según su alcance.

En relación con la topología de la red encontramos dos grupos: aquellas en que la red presenta una infraestructura formada por un número de estaciones inmóviles (nodos) a las que acceden los terminales, y aquellas en las que los propios terminales forman una red *ad–hoc*, siendo ellos mismos los nodos de esta.

Según su alcance, y variando este de menor a mayor, podemos dividir las redes inalámbricas en Redes Inalámbricas de Área Personal (Wireless Personal Area Network, WPAN), Redes Inalámbricas de Área Local (Wireless Local Area Network, WLAN) y Redes Inalámbricas de Área Amplia (Wireless Wide Area Network, WWAN). Está clasificación se emplea con frecuencia, por lo que veremos los tipos anteriores con algo más de detalle.

Una red WPAN tiene un alcance corto, de unos 10 metros, y utiliza una frecuencia que no requiere de licencia para operar. La mayoría de las redes de este tipo se basan en Bluetooth, y su velocidad de transmisión es de unos 0.5 Mbps.

Por su parte, una red WLAN tiene un alcance mayor, entre 10 y 100 metros, y su velocidad es muy superior, hasta los 100 Mbps. Utilizan también frecuencias sin necesidad de licencia. Las redes inalámbricas de este tipo surgen a partir de las redes locales no inalámbricas (LAN), principales pensadas para la transmisión de datos. Es por ello que esta tecnología está principalmente orientada a la transmisión de datos, y no ofrece soporte para voz como sucede con las redes WWAN.

Una red WWAN cubre una distancia de entre 100 metros y 30 kilómetros, y emplea una frecuencia no libre, es decir, una cuyo uso requiere la adquisición de una licencia. Originalmente este tipo de redes se pensaron para transmisión de voz, por lo que su velocidad es baja, 4,8 kbps. La evolución de estas redes para la transmisión de datos ha dado lugar a una segunda generación con mayores velocidades, como sucede con las redes de los sistemas GSM (Global System for Mobile) o GRPS (General Packet Radio Service), con velocidades de 9,6–14 kbps y 20–115 kbps respectivamente. Estas velocidades siguen siendo insuficientes para gran cantidad de aplicaciones, pero las redes de tercera generación, como el sistema UTMS (Universal Mobile Telecommunication System) europeo, pueden alcanzar tasas que permiten operar fluidamente del mismo modo que en una red local.

25.7. El *sofware* SIG móvil

Conocemos ya los elementos que integran el SIG móvil y las tecnologías implicadas tales como las redes inalámbricas y los métodos de posicionamiento. Es el momento de ver cómo el *software* SIG se adapta a estas circunstancias y cuáles son las características de las aplicaciones que vamos a encontrar sobre los dispositivos móviles.

Las diferencias entre los SIG de escritorio y los SIG sobre dispositivos móviles vienen motivadas fundamentalmente por dos razones: las capacidades limitadas de estos (que mencionamos al inicio del capítulo) y las funcionalidades extras que presentan (principalmente la capacidad de posicionamiento). De igual modo, el enfoque y el tipo de uso que se pretenda dar condicionan la forma de las aplicaciones, existiendo una gran diferencia entre las aplicaciones dirigidas a lo que denominábamos SIG en campo y aquellas orientadas a los servicios basados en localización.

Comenzando con las primeras, representan el *software* más similar a los SIG de escritorio, ya que las funcionalidades que resultan de interés son en buena medida aquellas que encontramos en estos. La lectura de datos y su representación son de nuevo los pilares fundamentales entre las capacidades que una aplicación para SIG en campo debe presentar, aunque tanto la edición como el análisis cobran relevancia y se implementan habitualmente para usos particulares. A su vez, tanto la lectura como la representación de datos son dos de las áreas en las que es más necesaria una adaptación debido a las limitaciones del dispositivo.

En el caso de la lectura de datos, la limitada capacidad de almacenamiento y, sobre todo, memoria y velocidad de proceso, plantean un problema a la hora de desarrollar un *software* que se comporte de manera similar a un SIG de escritorio. Aunque el desarrollo de ciertos tipos de dispositivos móviles es rápido y sus capacidades casi alcanzan en algunos casos a las de un ordenador de sobremesa, el manejo de datos voluminosos sigue estando restringido. Este tipo de datos, no obstante, no son necesarios con tanta frecuencia como en el trabajo clásico con un SIG de sobremesa y, dado que otro tipo de funcionalidades están más limitadas, el rango de actividades que se van a desarrollar con tales datos es más reducido, lo que simplifica el desarrollo de todo lo relativo a su acceso y manejo.

Aunque un SIG móvil era en su concepción inicial un elemento autónomo capaz de contener los datos necesarios para su funcionamiento e incluso incorporar nuevos datos mediante la creación *in situ* de estos, la aparición de las redes inalámbricas ha cambiado esta tendencia y ahora el desarrollo se enfoca hacia el consumo de datos externos a través de la red. Este planteamiento soluciona las dificultades que existen para la lectura de datos de gran volumen, ya que el dispositivo se convierte en un cliente y delega las tareas más costosas al servidor correspondiente.

En los dispositivos de mayor potencia, adecuados para un desarrollo profesional del SIG en campo y para la recogida de datos, el SIG conserva sus capacidades de acceder a datos locales, mientras que en otros menos potentes y especializados, tales como teléfonos móviles, se consumen exclusivamente datos remotos. Algunas aplicaciones con base SIG, tales como navegadores, pueden utilizar cartografía digital almacenada en el dispositivo, pero la aplicación como tal no permite la utilización de otros datos distintos o la lectura de diversos formatos, como sí sucede en un SIG de escritorio.

En lo referente a la representación, la principal diferencia que se ha de considerar a la hora de diseñar un SIG móvil es, como parece lógico, la reducida dimensión de las pantallas. Especialmente a la hora de visualizar datos y aplicar una simbología a estos, se ha de tener en cuenta que existe una limitación de tamaño y que no pueden aplicarse ideas idénticas a las que serían adecuadas para una pantalla de ordenador común, ya que, al trasladarlas a la del dispositivo móvil, puede obtenerse como resultado un mapa carente de utilidad que no transmite adecuadamente la información geográfica que contiene. Los conceptos de generalización cartográfica que mencionamos en el apartado 3.5 (por ejemplo, la exageración de elementos) han de tenerse muy presentes en la creación de un SIG móvil.

No solo en la forma de representación existen diferencias, sino también en las propias funcionalidades de visualización incorporadas en la aplicación. Esto está relacionado no únicamente con las limitaciones de la aplicación —podemos decir que, en general, el SIG sobre un dispositivo móvil es una versión más simplificada y menos compleja de un SIG de escritorio—, sino con las

necesidades que el usuario va a tener en este aspecto.

Por ejemplo, podemos asumir que un usuario de un SIG móvil va a requerir menos capacidades para establecer una representación particular de los datos espaciales, ya que el trabajo que realiza es menos exigente en ese sentido. Mientras que sobre un SIG de escritorio se elabora cartografía y se trabaja con múltiples capas y en contextos de trabajo muy distintos, un usuario de un SIG móvil emplea la representación visual de los datos como forma de navegación (de modo similar a como emplearía un mapa en papel), o como un apoyo para la edición o toma de datos. En el primer caso, en la representación debe primar la claridad, para facilitar la localización de aquello que busca. Aspectos relativos al análisis visual de la componente temática del dato geográfico no son relevantes, ya que es raro que el usuario efectúe ese tipo de operaciones. En el segundo caso, debe prevalecer la representación clara de aquello que se edita o de los elementos principales del entorno que van a servir de guía para la edición o creación de nuevos datos.

Aunque también los SIG móviles tienen parte del carácter generalista de los SIG de escritorio, su contexto está más acotado o, al menos, más limitado en cuanto a la extensión de las actividades que pueden llevarse a cabo y las necesidades que van a plantear. Por esta razón, sus funcionalidades, con la visualización en lugar predominante, también se encuentran limitadas.

Gracias al acceso a Internet que se mencionó anteriormente, no solo las tareas de acceso y procesado de datos se delegan en un servidor, sino también las relacionadas con la representación. Por eso, es más habitual que los SIG móviles actúen como clientes de servicios de mapas (es decir, de representaciones ya hechas y listas para visualizarse, como vimos en el apartado 24.5.1), y no como clientes de servicios más complejos en los cuales se obtienen los datos y después es la aplicación la que se encarga de formar la representación a partir de ellos.

Esto no quiere decir que este tipo de capacidades no se encuentren en los SIG móviles. De hecho, algunas aplicaciones SIG móviles permiten incluso que la edición de la cartografía sea también un servicio remoto, es decir, que cuando el usuario edite o añada nuevos elementos en su trabajo de campo, estos cambios no tengan lugar en los datos locales que existen en el dispositivo, sino que modifique los presentes en un repositorio remoto. Esta funcionalidad, poco frecuente incluso en los SIG de escritorio más completos, aparece en algunos SIG móviles. No obstante, las posibilidades de representación son menores en el SIG móvil, entendiéndose que no es necesario ofrecer capacidades avanzadas de este tipo.

A modo de ejemplo, y tras lo explicado hasta este punto, se muestra en la figura 25.3 el aspecto de una aplicación SIG móvil.

En los servicios basados en localización, todo lo anterior tiene lugar de un modo aún más patente, reduciéndose por lo general más aún las funcionalidades. El usuario tiene menos capacidad para «operar» con el dispositivo y con el *software*, y los servicios se diseñan para que sean sencillos de consumir. Los teléfonos móviles, que representan el dispositivo por excelencia para este tipo de aplicaciones, tienen capacidades más reducidas que otros de los adecuados para el SIG móvil, por lo que esta limitación de funcionalidades es también producto del dispositivo al que están orientadas mayoritariamente. La menor especialización de los usuarios influye también en que las aplicaciones presenten esas características.

La imagen 25.4 muestra dos ejemplos de aplicaciones para servicios basados en localización. Adviértase que estas no tienen necesariamente que guardar similitud con la idea clásica de un

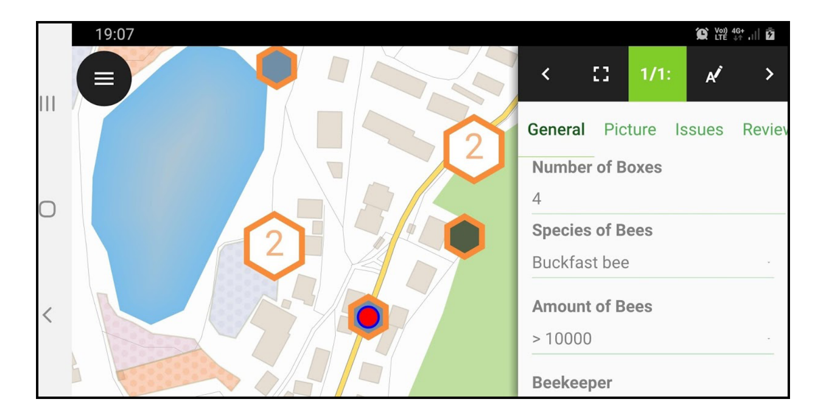


Figura 25.3: Aspecto de una aplicación SIG móvil (QField) (Cortesía OPENGIS.ch)

SIG, y que pueden no incluir ningún tipo de representación cartográfica. Es decir, que pueden proveer el servicio dando algún tipo de información geográfica (en el caso del ejemplo de la izquierda, se ofrecen mensajes de otros usuarios del servicio localizados en la misma zona) sin necesidad de mostrarla sobre un mapa. En el caso de la captura de pantalla mostrada en el lado derecho, la información sí aparece en un mapa, en el cual se muestran los contactos del usuario que se encuentran cercanos. El servicio en este caso es una forma particular de agenda de contactos que hace énfasis en algunos de ellos en función de su localización y la del usuario.

25.7.1. El contexto

Un hecho básico a considerar a la hora de diseñar *software* para un SIG móvil es que en este el *software* conoce dónde se encuentra el usuario, y el trabajo de dicho usuario normalmente se basa en emplear esa localización para realizar algún tipo de tarea. Aparece así un concepto que carece prácticamente de importancia en un SIG de escritorio que se ejecuta sobre una máquina inmóvil, pero que en el SIG móvil y en cualquier otra aplicación móvil resulta fundamental: el *contexto*.

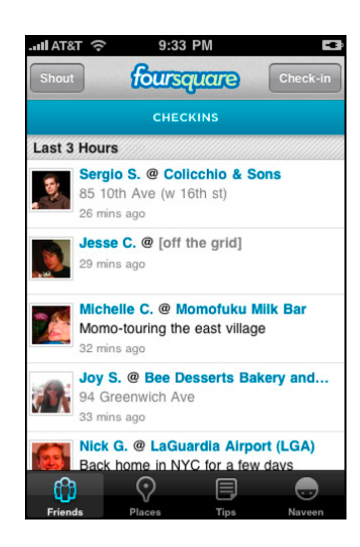
Entendemos por contexto toda aquella información que puede ser utilizada para caracterizar la situación de una entidad. Una entidad es una persona, lugar o objeto que se considera relevante para la interacción entre el usuario y la aplicación, pudiendo considerarse como entidad también a estos últimos [321].

Los factores implicados en definir un contexto son variados, pero pueden considerarse divididos en cuatro grupos fundamentales [322]:

- Contexto **espacial**. Caracterizado por dónde se encuentra el usuario.
- Contexto **social**. Caracterizado por quién es el usuario.
- Contexto **informacional**. Caracterizado por qué recursos se hallan cerca del usuario.
- Contexto **técnico**. Caracterizado por las características de la red y los dispositivos.

Si atendemos al caso particular de los servicios basados en mapas, [323] propone los tipos de contexto que se detallan a continuación:

- **Usuario**. La identidad del usuario permite considerar aspectos tales como su edad y sexo (las cuales condicionan inevitablemente sus intereses), sus preferencias personales (por ejemplo, el idioma que habla y en el que quiere recibir el



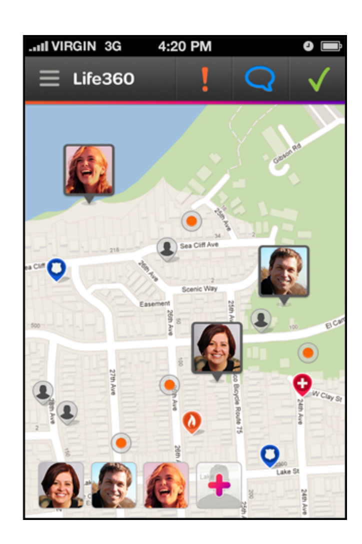


Figura 25.4: Dos servicios basados en localizacion sobre un teléfono móvil. En la parte izquierda, Foursquare [http://www.foursquare.com]. En la parte derecha, Life360 [http://www.life360.com].

servicio) o quiénes son su amistades si y desea contactar con ellas.

- **Localización**. El elemento de contexto más empleado, puede ser tanto absoluta (expresada mediante una coordenada georeferenciada) o relativa a algún otro elemento que forma a su vez parte del contexto.
- **Tiempo**. Puede considerarse a distintas escalas. Por ejemplo, la hora del día (de interés si se busca un establecimiento para indicar al usuario solo aquellos que estén abiertos en ese momento) o la estación del año (que condiciona las actividades que se pueden realizar, ya que muchas de ellas son estacionales).
- **Orientación**. Para saber hacia dónde se dirige el usuario y conocer, por ejemplo, qué tiene delante a la vista. También para servicios de navegación, para saber si el usuario sigue adecuadamente una ruta propuesta. Si el usuario se mueve, puede conocerse mediante el movimiento, pero en caso de estar parado requiere la presencia de elementos adicionales en el dispositivo.
- **Historial de navegación**. Permite crear un perfil del usuario y saber sus intereses en función de los lugares en los que ha estado.
- **Propósito de uso**. Viene definido por las actividades y objetivos del usuario, así como el papel que ejerce durante la utilización del dispositivo móvil. Los distintos tipos de usuarios tendrán diferentes necesidades en lo que respecta a la información, la presentación (por ejemplo, mapas con una representación más o menos técnica) o los modos de interacción con el dispositivo.
- **Situación cultural y social**. La situación de un usuario en este sentido se caracteriza por la proximidad a otros usuarios, su relación social y sus tareas colaborativas.
- **Entorno físico**. En este apartado se incluyen elementos como la iluminación existente o el ruido ambiente, que condicionan la interacción con el dispositivo y las capacidades del usuario de operar sobre él.
- **Propiedades del sistema**. Se incluyen aquí los aspectos relativos a la tecnología. Por ejemplo, si el dispositivo es en color o en blanco y

negro, si tiene teclado o pantalla táctil, o si la conexión a Internet es continua o intermitente.

Algunos de los anteriores puntos puede pensarse que no guardan una relación directa con los LBS y no han de ser exclusivos de estos. Es decir, que elementos como, por ejemplo, el tiempo, pueden ser tenidos en cuenta a la hora de proveer un servicio sin necesidad de que el dispositivo a través del que se realiza dicho servicio cuente con medios para establecer su posición. Un ordenador de sobremesa, por ejemplo, también dispone de información sobre el tiempo que puede considerarse. Aunque esto es cierto, la inclusión del contexto espacial añade relevancia a los otros elementos del contexto, ya que modifica en gran medida la labor del usuario y la naturaleza de su actividad sobre el dispositivo.

Si recurrimos al clásico ejemplo del cálculo de rutas, aunque el análisis llevado a cabo sea similar y requiera unos datos similares (punto de inicio, punto de destino y red de vías de comunicación), el hecho de que uno de dichos puntos (habitualmente el de salida) sea la coordenada actual del dispositivo modifica en gran medida muchos aspectos de esa operación. Al realizar un cálculo de rutas en un SIG de escritorio sobre un ordenador de sobremesa, lo normal es que este cálculo nos sirva para planificar un viaje futuro o para estimar el tiempo que, en algún momento dado, tardaremos en cubrir la distancia entre dos puntos.

Al contrario que en el caso de usar un dispositivo móvil como un navegador GPS, ese viaje por esa ruta no vamos a realizarlo inmediatamente, no tiene necesariamente que ser un trayecto cercano a nuestra posición actual e incluso no vamos a ser nosotros mismos quienes hagamos el recorrido. De este modo, el contexto temporal o el personal del usuario no tienen significado alguno. Podemos incluir esas variables explícitamente si el *software* así nos lo permite, pero no son una parte inherente al cálculo y que siempre sea tenida en cuenta. Por su parte, en un dispositivo móvil pueden incorporarse todos estos factores asumiendo que, en la mayoría de los casos, sí van a ser de importancia.

En resumen, que el hecho de que se trabaje sobre un dispositivo móvil y este permita conocer su posición añade significado a todas las clases de contexto.

El *software* debe diseñarse de forma que pueda responder a ese contexto y adaptarse a él. Las siguientes son las áreas principales en las que esa adaptación puede producirse [324]:

- **Información**. La información proporcionada a un usuario varía en función del contexto en que se encuentre. Por ejemplo una búsqueda de un determinado tipo de comercio puede restringirse a un radio de alcance desde su posición habitual, o bien filtrarse para informar solo de aquellos que ofrezcan algún producto o servicio que sea de interés para el usuario.
- **Tecnología**. Conociendo las características del dispositivo, puede establecerse la mejor forma de ofrecer un servicio. Si, por ejemplo, la pantalla del dispositivo es demasiado reducida, no será interesante hacerlo mediante imágenes de gran tamaño, cómo tampoco lo es proporcionar algún tipo de información sonora si el dispositivo no dispone de capacidades para reproducir sonido. Como ya vimos, una de las características del SIG móvil es la variedad de plataformas, por lo que la adaptación en este sentido es importante para poder satisfacer las necesidades de los usuarios con independencia de qué plataforma emplean.
- **Interfaz de usuario**. El servicio puede alterar la interfaz sobre la que opera el usuario. El ejemplo más clásico es el desplazamiento de un mapa a medida que este se mueve.
- **Presentación**. Si la información requiere ser representada, esto puede hacerse de diversas for-



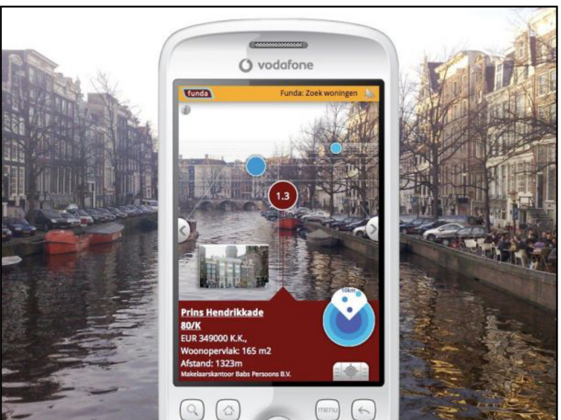


Figura 25.5: Dos ejemplos de realidad aumentada (cortesía de 5 Magazine)

mas en función del contexto. La simbología empleada se adapta, por ejemplo, a las preferencias del usuario (resaltando aquellos elementos que le resultan de mayor interés) o a la hora del día (señalando de algún modo explicito el hecho de que algunos elementos pueden no estar disponibles, tales como comercios si es de madrugada), entre otros factores.

La adaptación a un contexto dado puede ser mayor o menor en función de las propias características del servicio y de cómo este se plantee. En algunos casos puede llegar a ser muy intensa, tal y como sucede en la denominada *realidad aumentada*, donde la frontera entre la realidad y el dispositivo se difumina gracias a que aquella se «sumerge» en este y es ampliada. En la realidad aumentada, vemos en la pantalla de nuestro dispositivo imágenes del entorno en el que nos encontramos, pero complementadas con elementos adicionales tales como gráficos, vídeos o sonido. Estos elementos es posible incorporarlos gracias a que se conoce con exactitud el contexto, y esa información puede emplearse para buscar nueva información que añadir. La figura 25.5 muestra dos ejemplos muy ilustrativos de lo anterior.

25.8. Resumen

Los SIG móviles combinan las tecnologías SIG con los dispositivos móviles, el acceso inalámbrico a Internet y los sistemas de posicionamiento, para ofrecer una solución ventajosa para el desarrollo de trabajo de campo. De particular interés son los denominados Servicios Basados en Localización, donde un tercero ofrece servicios que dependen de la posición del dispositivo en cada momento. Otras de las tareas fundamentales del SIG móvil son la navegación o la captura de datos espaciales directamente en el dispositivo, las cuales son las principales en lo que hemos denominado SIG «en campo».

Para comprender el funcionamiento de las tecnologías implicadas en el SIG móvil, hemos analizando por separado los métodos de posicionamiento, las redes inalámbricas y las aplicaciones de *software*, cada una de las cuales desempeña un papel básico en definir las capacidades de un sistema SIG móvil.

Parte V

La visualización

Capítulo 26

Los SIG como herramientas de visualización

26.1. Introducción

Visualizar la información geográfica es una parte fundamental del trabajo con un SIG. Aunque no es un aspecto imprescindible, y es posible incluso encontrar SIG enfocados al análisis en los cuales no existe forma de visualizar la información con la que se trabaja, la gran mayoría de soluciones incluyen las funcionalidades de visualización como elemento básico, y estas resultan imprescindibles para muchos de sus de usuarios.

Como ya vimos en el capítulo dedicado a las herramientas de escritorio, dos son las tareas que un SIG debe permitir en lo que a visualización respecta: crear representaciones dentro del entorno mismo del SIG y generar representaciones autocontenidas que puedan imprimirse y den lugar a un documento cartográfico en sentido clásico. La representación en pantalla dentro del SIG puede guardar similitud con la idea clásica de mapa, o bien ser distinta, aprovechando elementos que no son habituales en esos mapas y que la tecnología del SIG sí que permite.

En ambos casos, lo más relevante de cara a los conocimientos que el usuario del SIG debe tener en cuanto a visualización es la capacidad de convertir los datos en elementos visuales útiles y con significado, con independencia de que estos vayan a representarse y usarse en pantalla durante una sesión de trabajo, o bien vayan a imprimirse en papel para su uso posterior en ese soporte. Este es el objetivo de esta parte del libro: proporcionar las ideas fundamentales para que el usuario de SIG logre las mejores representaciones visuales durante su trabajo con el SIG. Para ello, lo primero es conocer qué nos ofrece un SIG como herramienta de visualización y qué podemos esperar de él.

26.2. Particularidades del SIG como herramienta de visualización

Como herramienta de visualización, el SIG tiene sus particularidades, las cuales deben unirse a las propias de los modelos de almacenamiento que empleamos para recoger la información geográfica a visualizar. Esto hace que el trabajo de generar una representación visual de una determinada información geográfica no sea igual en el caso de realizarse mediante un SIG que cuando se lleva a cabo en base a la labor clásica del cartógrafo. Trabajar en un SIG añade, entre otros elementos, el hecho de que la información se encuentra almacenada según un modelo dado (ráster o vectorial). Si esta distinción implica, como ya sabemos, notables diferencias a la hora de analizar esa información u optimizar el acceso a los datos que la contienen, no es menos cierto que también va a conllevar un enfoque distinto a la hora de visualizar unos u otros tipo de datos.

Para el cartógrafo en su concepto clásico, esta distinción no existe. Indirectamente, sí puede asumirse que existe algo similar, ya que el cartógrafo ha de conocer la naturaleza de las variables que representa, y sabemos que esta naturaleza se encuentra muy ligada al modelo a escoger para representarla (por ejemplo, sabemos que variables continuas como la elevación se analizan mejor si se almacenan según el modelo ráster, aunque ello no implica que no puedan almacenarse de un modo distinto y ello no tenga interés). Sin embargo, no existe una división formal explicita tal como sucede en el caso del SIG.

Otra de las diferencias a la hora de representar la información geográfica en un SIG deriva del propio objetivo que dicha representación tiene. La labor del cartógrafo tiene como fin primordial el crear un elemento visual que transmita la información geográfica. El cartógrafo, por lo general, no es un usuario de la cartografía, sino un productor de esta para su uso por terceros. El usuario de SIG, sin embargo, puede crear cartografía para otros pero, en la mayoría de los casos, la crea para sí mismo para poder emplearla como una herramienta más a la hora de desarrollar su trabajo con el SIG. Por esta razón, la representación visual que se produce con un SIG puede tener un carácter general y estar pensada para ser empleada en ámbitos diversos, pero también puede tener una funcionalidad muy clara dentro de un campo de aplicación dado, o incluso dentro exclusivamente de un proyecto concreto. Este hecho puede relajar las exigencias que se presentan al generar una representación cartográfica en un SIG, pero al mismo tiempo también ofrece la posibilidad de enfocar el esfuerzo de visualización de forma más particular. Es decir, de considerar el contexto de ese ámbito de utilización para lograr una representación más eficaz dentro de ese entorno particular.

Por último, cuando pensamos en un mapa tradicional es difícil advertir que se trata de un elemento visual creado a partir de otro no visual. Es decir, un mapa es un elemento gráfico desarrollado a partir de unos ciertos datos. Los datos en sí no pueden «verse», pero son los que posibilitan la creación del mapa. Resulta más sencillo pensar que el cartógrafo simplemente plasma la realidad del terreno (que podemos ver con nuestros propios ojos sin más que ir a la zona representada por el mapa), que pensar que está convirtiendo datos en elementos gráficos tales como líneas o puntos. Sin embargo, y aunque no utilicemos un SIG, esos datos existen, ya que el cartógrafo pone sobre el papel las medidas (datos numéricos) tomados por los técnicos en campo, ya sean estas provenientes de algún sensor o el resultado de un levantamiento topográfico, entre otros orígenes posibles.

El beneficiario de esos datos es el usuario del mapa, que recibe la representación visual de éstos, y es esta visualización la que le transmite, en la medida de lo posible y en función de su propia calidad como elemento de comunicación, la información geográfica subyacente. El usuario de un SIG, salvo que utilice una imagen (una fotografía aérea, un mapa escaneado o un mapa ya preparado a través de un servicio de mapas), no recibe elemento visual alguno, sino tan solo datos numéricos que son, eso sí, muy susceptibles de ser visualizados.

En relación con esto, un SIG está pensado para satisfacer dos necesidades fundamentales. Por una parte, la creación de cartografía a partir de los datos, del mismo modo que el cartógrafo utiliza otro tipo de herramientas para elaborar mapas a partir de los datos topográficos o temáticos de los que dispone. Por otra, y para el usuario cuyo fin último no es la elaboración de cartografía, visualizar de la mejor forma posible los datos con los que trabaja, para que esta visualización aporte

valor añadido a los datos de cara al desarrollo de la labor de ese usuario. Ambos enfoques coexisten en un SIG y están orientados en cualquier caso a extraer de los datos la mayor información posible de forma visual.

En definitiva, debemos tener siempre presente que en un SIG la información geográfica no es un elemento visual, ya que llega a nosotros convertida en última instancia en algo puramente numérico, apto para ser procesado de un modo u otro por el ordenador en el que ejecutamos el SIG. Somos nosotros, a través del SIG, quienes la dotamos de un aspecto visual.

En otras palabras, en un mapa clásico la tarea del cartógrafo (que es quien prepara la información geográfica) es hacer que sea lo más fácil posible de interpretar para el usuario de ese mapa. En el SIG existe también alguien que prepara los datos (por ejemplo, un técnico que comprueba la calidad de un MDE y lo almacena en un formato dado), pero su objetivo es facilitar su interpretación y uso al ordenador (o, más concretamente, al SIG). La visualización, por lo general, y salvo que en esa preparación se añadan elementos adicionales que complementen al dato en sí, queda en manos del usuario del dato. Es por esta razón que una parte como esta resulta fundamental en un libro de SIG, ya que el usuario de SIG necesita conocer cómo emplear el SIG para visualizar la información con la que trabaja.

26.3. La visualización científica y los SIG

La visualización de datos no es algo exclusivo de los SIG como aplicaciones informáticas, y en absoluto se trata de algo nuevo relacionado con los ordenadores y sus capacidades de representación. La creación de gráficas y diagramas es una realidad desde mucho antes que aparecieran los ordenadores, y estas son una herramienta fundamental en el ámbito científico. Visualizar series de datos sencillos mediante la representación de éstos ayuda a comprender su naturaleza y constituye un útil de gran potencia a pesar de su aparente simplicidad.

Visualizando un dato cualquiera se obtiene una densidad de información mucho mayor que si ese mismo dato se representa numérica o textualmente. Se estima que aproximadamente el 50 % de las neuronas están dedicadas a la visualización. Como reza la sabiduria popular, «una imagen vale más que mil palabras», y esta es una verdad que cobra pleno sentido dentro del campo de las ciencias.

Se puede pensar que una representación simple tal como un diagrama de barras o uno de dispersión está muy alejado del tipo de representación compleja que un mapa supone, y que, por tanto, también es muy distinta de la representación que tiene lugar en un SIG. Analizándolo con un poco más de detalle vemos, sin embargo, que la diferencia no es tal y existen muchas similitudes y relaciones.

En primer lugar, estas representaciones pueden aplicarse a la componente temática de los datos espaciales y permitir el análisis de esta. Prescindiendo de la componente espacial, la componente temática es susceptible de ser analizada mediante cualquiera de las herramientas habituales de la estadística descriptiva, entre ellas las del análisis exploratorio de datos tales como las gráficas y diagramas antes mencionados.

En segundo lugar, existen en la actualidad otras herramientas de visualización de datos más avanzadas, producto del avance tecnológico de los últimos tiempos, el mismo que ha propiciado el salto de la cartografía clásica a la cartografía digital y al campo de los SIG. Si el volumen de información y sus características ha variado sensiblemente en lo que al ámbito geográfico respecta, otras áreas no han sido ajenas a transformaciones

similares, lo cual ha tenido como consecuencia el desarrollo de nuevas ideas para poder visualizar esa información y poder aprovechar sobre ella las capacidades de percepción y análisis visual de que disponemos. El desarrollo en este sentido es tal que constituye en la actualidad una rama de la ciencia propia: la *visualización científica*.

Los conceptos de la visualización científica pueden ser aprovechados por los SIG, que aproximan de ese modo sus funcionalidades a las de las aplicaciones de visualización genérica de datos. En algunos casos, las diferencias son meramente formales y debidas a los enfoques tradicionales que se vienen empleando en estos campos, y la integración entre ambos es posible al menos en lo que a sus conceptos y fundamentos respecta.

Consideremos por ejemplo, las representaciones de la figura 26.1. La de la izquierda se ha producido a partir de datos obtenidos en un túnel de viento y muestra las presiones ejercidas por el aire sobre un ciclista, así como las líneas de flujo que se forman. La de la derecha representa la actividad cerebral en un ratón tras un estímulo, y se ha creado en base a los datos proporcionados por un tomógrafo. Salvando las diferencias en cuanto al campo de la ciencia del que provienen, ambas representaciones guardan muchas semejanzas con, por ejemplo, las obtenidas a partir de un MDE, en las que habitualmente se emplea una paleta de colores similar para visualizar los valores de las distintas celdas. Puedes ir al capítulo 14 para encontrar un buen número de ellas y comprobar por ti mismo esa similitud.

Aunque puedan parecer algo muy alejado de lo que entendemos por SIG y del trabajo que desarrollamos con uno de ellos, lo cierto es que representaciones así podrían perfectamente formar parte de un proyecto SIG, al menos en teoría. Si pensamos en la primera de ellas, la del ciclista, no es raro en la actualidad que un SIG 3D permita cargar modelos tridimensionales tales como edificios o árboles, por poner dos ejemplos (veremos esto con detalle algo más adelante en esta misma parte del libro). De este modo, no es tan descabellado pensar en disponer en un SIG de los datos de la forma de ese ciclista, datos que, por otra parte, son de tipo espacial y encajan perfectamente en lo que un SIG esta diseñado para manejar.

De hecho, el modelo que ha servido para calcular esos datos de presión podría aplicarse mediante las capacidades de modelización de un SIG, y podría estudiarse un supuesto en el que se conocieran los datos de viento de una determinada zona. Es decir, situar al ciclista en una calle dada y con unas condiciones concretas y efectuar el cálculo que nos llevaría a unos datos similares a los representados en la imagen. Siendo posible realizar ese cálculo en un SIG, visualizar esos datos resultantes a través de una representación como la mostrada es, sin embargo, algo que no resulta aún posible en un SIG, y es necesario el concurso de una aplicación especializada de visualización científica.

Así pues, las imágenes de la figura 26.1 no han sido creadas con un SIG, sino con sendas aplicaciones de visualización científica de ese tipo. Estas aplicaciones presentan funcionalidades distintas a las que tiene un SIG, habitualmente más avanzadas y con un mayor grado de interactividad. Asimismo, están pensadas para la representación de datos multidimensionales, algo que no sucede con los SIG [325]. La diferencia principal estriba en que, mientras que la visualización en el SIG complementa al análisis y a otras operaciones sobre los datos, en la visualización científica esta *es* el análisis, y el objetivo único de la visualización es facilitar el análisis visual de los datos. Este es el motivo por el que aparecen funciones avanzadas de tipo interactivo que permiten al usuario «jugar» con los datos, alte-

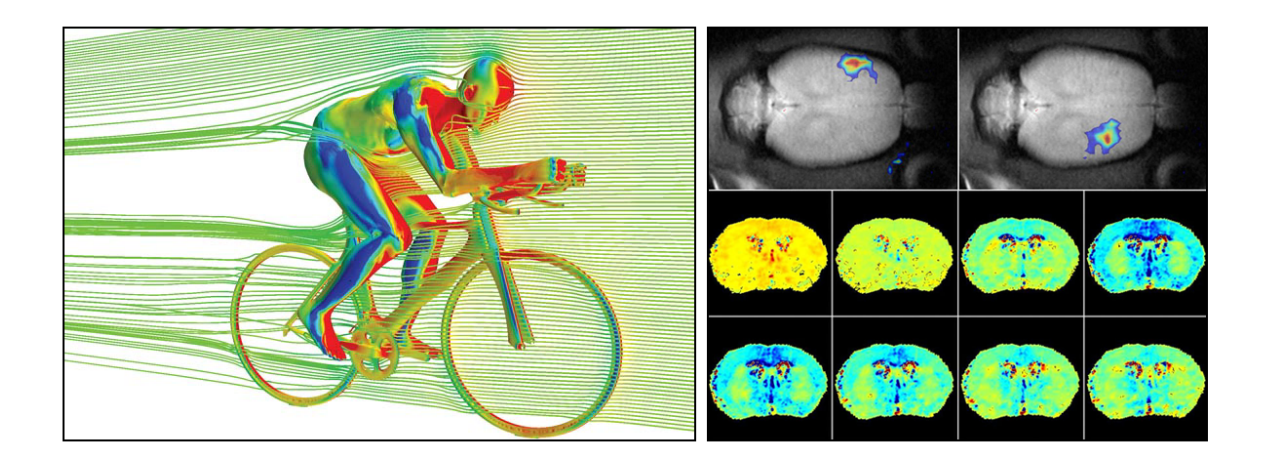


Figura 26.1: Dos ejemplos de visualización científica.

rando su representación para hacer más explicita la información que contienen.

Si estas funcionalidades avanzadas no aparecen en los SIG en la actualidad, esto no obedece a una imposibilidad técnica o a que carezca de sentido implementarlas, sino más bien al enfoque predominante en el diseño de los SIG, que en lo que a visualización respecta se asemeja mucho aún a la cartografía clásica. Aunque los SIG 3D van ganando terreno, la idea clásica de visualización en un SIG hereda directamente del mapa tradicional, y se constituye en muchos casos como una mera herramienta para crear este, sin considerar que puede ser posible la creación de otro tipo de representaciones.

Las limitaciones en cuanto a visualización también se deben en parte a las limitaciones en los datos, ya que un SIG no es de momento la herramienta ideal para el manejo de datos multidimensionales, a pesar de que estos abundan en el ámbito geográfico. Hemos estudiado mucho acerca de los datos espaciales en este libro, y la mayor parte de cuanto hemos visto se basa en el uso de geometrías planas o, en todo caso, tridimensionales, siendo extraño el trabajo con otros datos, al menos en los SIG de uso genérico. Existen, por ejemplo, modelos para mallas de datos multidimensionales, pero las capas ráster tal y como las hemos estudiado son puramente bidimensionales. Mientras haya carencias en los modelos de datos y en la concepción del dato geográfico, es lógico entender que las capacidades de visualización de los SIG también presenten deficiencias a la hora de trabajar ciertos tipos de datos

El uso combinado de aplicaciones para visualización científica y SIG es la solución actual a determinados problemas de visualización que exceden las capacidades habituales de estos últimos. En este sentido, se han producido acercamientos entre ambos tipos de aplicaciones para tratar de conseguir que esta combinación no se lleve a cabo tan solo mediante una mera compartición de datos (uso de formatos comunes que permiten «pasar» los datos de una aplicación a otra), sino que exista una verdadera integración que reduzca la redundancia de funcionalidades y maximice las posibilidades. Por el momento, la plena integración dista mucho de ser una realidad, por lo que debe recurrirse a la utilización conjunta de

una u otra manera. En [326] puede encontrarse este tema desarrollado con más profundidad.

Aunque en los SIG faltan muchos de los elementos y de las capacidades de las aplicaciones de visualización científica, algunas ideas de esta sí que aparecen en ellos, y en su conjunto amplían la potencialidad del mapa como metáfora de una realidad que se representa. La más básica de todas ellas es la interactividad que permiten las herramientas de navegación. Aunque lejanas de lo que podemos encontrar en aplicaciones de visualización científica especializadas, ofrecen un respuesta por parte del mapa a las acciones de quien lo utiliza. Frente al carácter pasivo del mapa impreso, las representaciones dentro de un SIG son activas.

Otros elementos menos frecuentes son la incorporación de animaciones y la visualización tridimensional. En el capítulo 29 veremos más acerca de ambas.

Este nuevo enfoque que se produce en el ámbito cartográfico al incorporar parte de las ideas de la visualización científica se conoce como *geovisualización*, y conforma una rama de esta última dedicada al caso particular de visualizar la información geográfica. Una forma muy gráfica de ver la diferencia entre el documento cartográfico clásico y la geovisualización que se produce dentro de un SIG es mediante el denominado *Cubo cartográfico* [327] (Figura 26.2).

El cubo cartográfico contiene tres ejes, en los cuales se representan el grado de interactividad, el objetivo principal de la representación y la audiencia a la que esta se dirige. La cartografía clásica y la geovisualización se sitúan en vértices opuestos, ya que presentan características distintas en estos tres conceptos. El mapa clásico esta pensado para presentar una información de la que ya se dispone, pero no es una herramienta para descubrir nueva información. La geovisualización, por el contrario, con la posibilidad que ofrece al usuario de «explorar» los datos, puede servir para extraer información que no se conocía de antemano a la hora de crear la representación. La interactividad es alta en la geovisualización y baja en el mapa clásico, como ya hemos visto. Por último, la audiencia en la geovisualización es privada, entendiéndose con esto no que existan restricciones para su acceso, sino que en su mayoría son representaciones fugaces que cambian según el usuario interactúa con el *software*, y por tanto lo normal es que solo sea ese usuario quien las disfrute, no teniendo un carácter persistente como el mapa impreso.

26.4. Los SIG frente a las aplicaciones de diseño

Pese a que, como acabamos de ver, la visualización en un SIG va mucho más allá del mapa tradicional, resulta indudable que la creación de este es una tarea fundamental y que los SIG han de responder a esa necesidad como herramientas primordiales para el cartógrafo y el diseñador. A pesar de ello, las necesidades del cartógrafo van a menudo más allá de los que un SIG puede ofrecer, siendo necesario recurrir a programas de diseño del mismo modo que sucede con las aplicaciones de visualización científica. Esto es así, principalmente, debido a que la labor del cartógrafo contiene un elemento artístico (que es, a su vez, puramente visual) que los SIG no están preparados para manejar. El SIG es una herramienta demasiado «estricta» en este sentido, ya que realiza una representación de los datos donde prima la exactitud y la corrección, sin dejar lugar para licencias que, si bien mejorarán la calidad del mapa como medio de transmisión de información, suponen un elemento fuera de la ortodoxia del SIG.

Asimismo, el SIG carece de funcionalidades para trabajar sobre los elementos artísticos y no

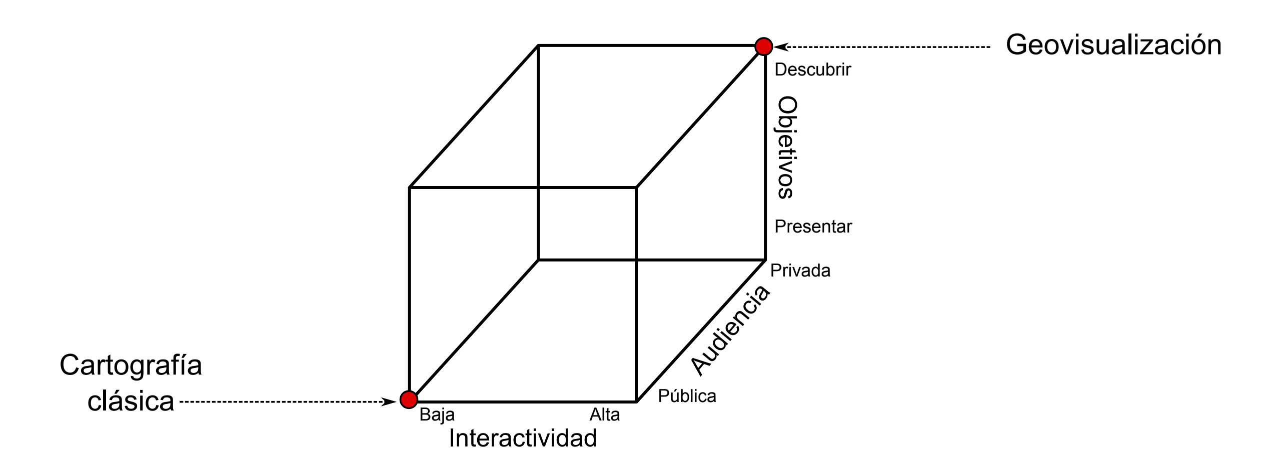


Figura 26.2: El *cubo cartográfico.*

cartográficos del mapa, o para editar libremente estos últimos más allá del propio rigor cartográfico. Y si estas funcionalidades aparecen, no suelen tener la suficiente potencia para un uso profesional.

La solución es hacer uso de aplicaciones de diseño que no tienen en consideración el significado de los elementos gráficos y no plantean restricciones a su modificación. Esto puede llevarse a cabo operando con el SIG para crear una primera representación, que luego se edita en un programa de diseño gráfico para retocar aquellos elementos que puedan mejorarse mediante el buen hacer del cartógrafo experimentado. En particular, el uso de *software* de ilustración vectorial es la opción más adecuada para la elaboración de mapas. Este planteamiento supone, sin embargo, una integración muy débil y que presenta numerosos inconvenientes, entre los cuales cabe citar los siguientes:

- **Incapacidad de la aplicación de diseño para analizar los datos**. La representación puede hacerse de forma completamente manual, creando cada uno de sus elementos y definiendo sus características sin la ayuda de ninguna rutina, pero también puede llevarse a cabo haciendo uso de alguna funcionalidad suplementaria. Por ejemplo, para establecer los colores de los distintos polígonos de una capa puede usarse el valor de uno sus atributos y establecer una rampa de colores en función de este. El SIG puede hacer esto automáticamente, pero una aplicación de diseño, puesto que no puede interpretar esos atributos y carece de esa funcionalidad, requerirá que el cartógrafo lleve a cabo esa asignación de colores de modo manual.
- **Dificultad de actualización**. Al no estar la representación sincronizada con la base de datos, las modificaciones en esta no le afectan, y es necesario rehacer los mapas cada vez que los datos cambien, ya que esa actualización no se produce de forma automática.
- **Nula o muy limitada capacidad de automatización de tareas**. Un SIG puede automatizar tareas tales como la subdivisión de un mapa en submapas menores (véase la imagen 23.4) o la producción de mapas sobre un conjunto de capas. Por ejemplo, podemos «mostrarle» al SIG cómo queremos el diseño del mapa de una variable dada y que él se encargue de generar los

mapas de ese modo para otra serie de variables recogidas en otras tantas capas en nuestra base de datos. Puesto que la aplicación de diseño gráfico no puede por sí misma acceder a esa base de datos, esta automatización no es posible en caso de crear cartografía con ella.

- **Mayor posibilidad de introducir errores cartográficos**. La permisividad de una aplicación de diseño gráfico es un arma de doble filo. Por un lado, permite al cartógrafo tomarse ciertas licencias cuando ello resulta necesario, pero también cuando no es correcto hacerlo. La aplicación no entiende, por ejemplo, que la orientación del mapa no debe variar si no lo hace también la rosa de los vientos, o que el *canevás* (la rejilla que acompaña al mapa) debe estar correctamente situado, y permite que se introduzcan errores que en un SIG se encuentran completamente controlados.

Al contrario de lo que sucedía con las herramientas de visualización científica, los SIG sí que van progresivamente incorporando la ideas de estas aplicaciones de diseño gráfico, permitiendo cada vez más la labor artística del cartógrafo y adaptándose a sus necesidades igual que se adaptan a las de otros usuarios con requerimientos distintos de visualización. Aún así, este tipo de capacidades deben considerarse como algo avanzado que pocos SIG incorporan, ya que la mayoría de ellos se centran en la visualización dentro de su propio entorno y solo permiten la elaboración de cartografía rudimentaria o, al menos, lejos de los estándares de la producción cartográfica clásica.

26.5. Resumen

La visualización es parte vital de los SIG, y por ello estos disponen de abundantes funcionalidades para la representar la información geográfica. Existen importantes diferencias entre la creación de una representación dentro de un SIG y la labor tradicional del cartógrafo. Desde el punto de vista conceptual, una diferencia fundamental es el hecho de que el usuario de la información geográfica en un SIG no la recibe en un formato visual, sino como meros datos numéricos, y ha de procurarse por sí mismo esa representación visual.

La visualización de datos es en la actualidad un apartado de gran importancia no solo en el campo del SIG, sino en todo el ámbito científico en general. Las aplicaciones existentes para la visualización de datos de diversa índole superan en muchas ocasiones a los SIG en cuanto a sus capacidades, especialmente en el manejo de datos multidimensionales y la interactividad entre el usuario y la representación. El uso conjunto de estas aplicaciones y los SIG amplia las posibilidades de estos.

Otras aplicaciones que complementan a los SIG en lo que a la producción de cartografía respecta son las empleadas en el diseño gráfico, cuyas capacidades también se van progresivamente incorporando a las aplicaciones SIG.

Capítulo 27

Conceptos básicos de visualización y representación

27.1. Introducción

Cuando visualizamos cualquier tipo de información geográfica, ya sea a través de un mapa clásico o de algún elemento gráfico en la pantalla de un ordenador, estamos utilizando un lenguaje visual para transmitirla. Del mismo modo que al hablar empleamos un lenguaje oral y al escribir un lenguaje escrito, siempre que plasmemos la información geográfica en una serie de elementos visuales estaremos empleando este lenguaje visual.

Existen muchas similitudes entre el lenguaje visual y el lenguaje que utilizamos cada día para comunicarnos, comunes todas ellas a cualquier forma de lenguaje. Por una parte, disponemos de una serie de elementos básicos que podemos usar, que son como las palabras con las que formamos frases y expresamos ideas. Estas se combinan de acuerdo con unas normas y siguiendo esquemas definidos que conocen tanto el creador del mensaje como el receptor, y sin los cuales no seria posible establecer la comunicación. Por otra, el conocimiento y manejo adecuado de todos estos elementos define nuestra capacidad de emplear el lenguaje y expresar correctamente aquello que queremos transmitir.

Al igual que en el lenguaje hablado, y por su carácter simbólico, el lenguaje visual implica la existencia de unas limitaciones. Es decir, no podemos expresar todo aquello que tratamos de representar, y un mapa nunca puede contener y transmitir fielmente toda la realidad de una zona o de un fenómeno espacial dado. Sin embargo, un correcto uso del lenguaje permite comunicar gran cantidad de información y hacer de este una herramienta valiosa, más allá de sus limitaciones, o incluso aprovechando estas para su propio beneficio.

El estudio de los signos de un lenguaje constituye lo que se conoce como *semiología*. En el caso de los elementos del lenguaje visual, encontramos una *semiología gráfica*, tal y como la definió el cartógrafo francés Jacques Bertin, pionero en este campo [328]. Esta semiología trata los signos del lenguaje visual y la gramática de estos, definiendo una lingüística visual que nos ayuda a comprender cómo una representación gráfica dada cumple su propósito de transmitir la información en base a la cual se crea.

Detallando las ideas de Bertin, en este capítulo veremos algunos aspectos fundamentales acerca del lenguaje visual que nos permitirán conocer sus propiedades y la forma en que sus elementos pueden emplearse de forma efectiva para la comunicación. Aplicando estos al caso particular de representar y visualizar información cartográfica, expandiremos este en el próximo capítulo para obtener un lenguaje cartográfico, fundamental para la creación de mapas.

27.2. Las variables visuales

Cada lenguaje tiene sus propiedades particulares y permite expresar unas u otras ideas de distintas maneras. Por ejemplo, podemos plasmar la música en una partitura, utilizando un lenguaje de signos musicales. Este lenguaje musical[1] permite recoger y transmitir una canción a través de una partitura, expresando mediante un conjunto de símbolos las distintas notas que la componen, su duración o los elementos expresivos que deben incorporarse en la interpretación de esta. Un músico que conozca este lenguaje puede interpretar una pieza gracias a que la música le llega a través de esos símbolos, siendo la partitura el medio de comunicación entre el interprete y el compositor o quien haya transcrito dicha pieza.

Aunque dos personas conozcan a la perfección el lenguaje musical, no podrán, sin embargo, transmitirse mediante sus símbolos y sus reglas algo como una formula matemática o un poema. El lenguaje matemático o el lenguaje oral son los adecuados para transmitir este tipo de mensajes, pero no el lenguaje musical, que tiene limitaciones en ese sentido.

Puesto que nuestro objetivo a lo largo de los capítulos de esta parte del libro es ser capaces de crear mapas y otros elementos visuales que transmitan la información geográfica, debemos estudiar qué clase de información vamos a transmitir y, sobre todo, qué nos permite transmitir el lenguaje visual. Del mismo modo que sabemos que los símbolos de nuestro lenguaje musical (pentagrama, figuras, etc.) no son capaces de transmitir una formula matemática, debemos ver si los elementos del lenguaje visual van a ser capaces de, por ejemplo, transmitir el patrón de distribución de un fenómeno en el espacio, las diferencias entre dos zonas distintas o la relación entre los valores de una variable en dos puntos. Además, debemos ver cómo emplearlos para que esa información se transmita de la mejor manera posible, ya que existen diversas propiedades de los elementos visuales que podemos emplear, siendo más adecuadas unas u otras según sea la circunstancia.

Estas propiedades conforman lo que se conoce como *variables visuales*, y se aplican a los elementos básicos de la representación, que son aquellos objetos geométricos de que se compone esta. Las variables visuales permiten diferenciar unos de otros y asignarles unas ciertas características, susceptibles a su vez de ser interpretadas junto al propio significado que el objeto pueda tener.

Dados dos elementos, estos pueden diferenciarse por las siguientes variables, que aparecen representadas en la figura 27.1:

- Posición
- Forma
- Color
- Tamaño
- Textura
- Orientación

Todas ellas constituyen las variables visuales, que estudiaremos seguidamente en detalle. El color, como explicaremos, se divide en dos variables visuales independientes: valor y tono.

Las variables visuales se aplican de forma distinta en función del tipo de elemento que queramos simbolizar, por lo que detallaremos su uso para las tres clases de símbolos que podemos incorporar en un mapa: puntuales, lineales y de superficie.

[1] Entiéndase que hablamos aquí de ese lenguaje de signos sobre un papel y no del lenguaje musical relativo a una forma de expresión a través del ritmo, el tono, etc.

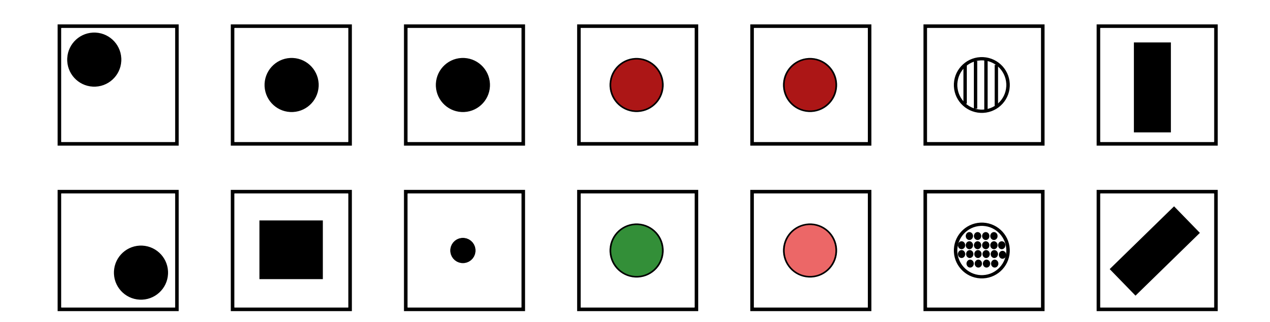


Figura 27.1: Ejemplo de uso de las distintas variables visuales. De izquierda a derecha: posición, forma, tamaño, tono, valor, textura, y orientación

27.2.1. Posición

La posición constituye un caso particular de variable visual a la hora de emplearla en la creación de cartografía, ya que viene fuertemente condicionada por el hecho de que todo aquello que representamos tiene una posición en el espacio y, por tanto, ha de tener una posición concreta en el mapa. Mientras que en cualquier otro tipo de gráfico la posición puede modificarse a voluntad para transmitir algún tipo de información, tal y como haremos con las restantes variables visuales, en el caso de un mapa la posición ya está asociada a una información que ha de transmitir: la información sobre la posición real en el espacio geográfico de aquel objeto que se simboliza.

Aunque el cartógrafo puede en determinadas ocasiones variar la posición de algunos elementos (por ejemplo, para mejorar la legibilidad del mapa), siempre está supeditado a la corrección cartográfica, y no posee libertad para alterar esta de cualquier modo. Por ello, el uso de la posición como variable visual está muy restringido en el caso de un mapa, y no se emplea. Su escasa aplicación en ese sentido queda patente en el hecho de que en algunos textos no se menciona junto a las restantes variables visuales, detallándose por separado como un elemento distinto.

27.2.2. Forma

La forma viene definida por el perímetro exterior del objeto. Esto no implica que únicamente se pueda aplicar la forma a símbolos de superficie, ni tampoco que se debe tratar de un perímetro cerrado como el de una forma poligonal.

La forma se aplica fundamentalmente a los símbolos puntuales, situando un símbolo de una forma dada sobre las coordenadas exactas del punto a representar. Su aplicación a símbolos lineales es difícil y no se da, mientras que en el caso de aplicarse sobre símbolos de superficie requiere la alteración de los polígonos representados, dando lugar a una representación imprecisa, al menos en lo que al contorno del polígono respecta. Esto se produce únicamente en el caso de los denominados *cartogramas*, un tipo particular de mapas que veremos en el próximo capítulo.

27.2.3. Color

La variable color es la más importante de todas las variables visuales, y la que a su vez requiere un grado mayor de detalle en su exposición, debido a la que complejidad que presenta y a las posibilidades que ofrece.

Existen muchas formas de representar y crear un color, a través de los denominados *espacios*

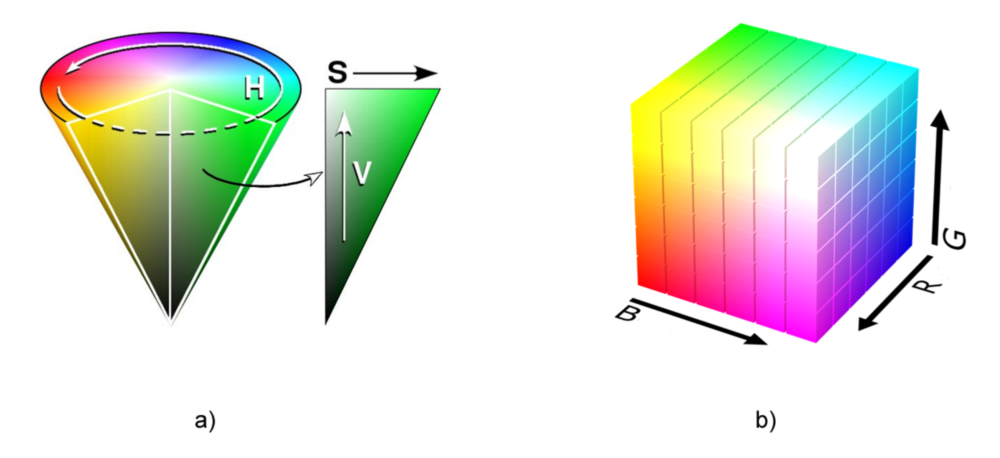


Figura 27.2: Espacios de color HSV (a) y RGB (b) (adaptado de Wikipedia).

de color. Ya conocemos el espacio de color RGB (figura 27.2b), que veíamos que era el habitual en el caso de imágenes, tanto para recoger la información como para representarla.

De cara a su uso como variable visual en el contexto de este capítulo, resulta de especial interés el uso del espacio de color HSV, en el cual un color se define mediante un espacio de coordenadas cilíndrico, según lo mostrado en la figura 27.2a.

Tres son las componentes de un color en este espacio de color —que es al que nos referiremos para el contenido de este capítulo—, las cuales establecen sus coordenadas en el cilindro: tono (H, de *Hue* en su denominación inglesa), valor (V) y saturación (S).

El tono es lo que en el lenguaje común denominaríamos color, es decir el nombre del color, por ejemplo verde, rojo o amarillo. Está relacionado con la longitud de onda de la luz, y distintas longitudes de onda producen un efecto perceptivo distinto, haciendo que distingamos así los diferentes colores. En el cilindro del espacio de color HSV, el tono viene marcado por el ángulo del vector definido por la posición del color y el eje central, sobre el plano perpendicular a dicho eje.

El tono puede verse alterado por los tonos del entorno, especialmente en símbolos de pequeño tamaño. Aunque es una variable para la que la percepción humana tiene gran sensibilidad, en los símbolos pequeños puede ser difícil de identificar y puede producirse una falsa percepción si comparten espacio con otras más grandes de un tono distinto. Por ejemplo, al trazar una línea con un grosor fino que atraviesa una serie de polígonos de distintos colores, el tono de esta se percibirá como distinto en cada uno de esos polígonos por el efecto que sus colores causan como colores de fondo.

Por su parte, el valor indica la claridad del color. Un tono azul puede ser más claro o más oscuro sin dejar de ser azul. Esa variación que se produce es una variación del valor del color. En el caso de usar una tinta de un color dado, la mezcla de esta con una pintura blanca produce una disminución del valor, aclarándose progresivamente según añadimos más de esta última en la mezcla. A la hora de imprimir, se hace uso de tramas más o menos densas para modificar el valor, sin modificar así la tinta. Según el espacio

en blanco que se deja entre los puntos de tinta impresos, se consigue la apariencia de un color de mayor o menor valor. El valor se define en el cilindro de coordenadas como la altura del color sobre el eje central.

La capacidad de diferenciar dos símbolos con valor distinto varía en función del tipo de símbolo. Así, es mayor en el caso de símbolos de superficie, mientras que en el caso de símbolos puntuales y lineales está relacionada con el tamaño. Si el punto es muy pequeño o la línea muy delgada, es más difícil apreciar el valor y, por tanto, comparar este con otro o extraer la información que mediante esa variable visual se intenta transmitir.

La saturación, por último, expresa la pureza relativa del color. Depende del número de distintas longitudes de onda que aparecen en un color dado. A medida que disminuye la saturación, el color va pareciendo más grisáceo, y el número de longitudes de onda es mayor. En el cilindro del espacio de color, queda definido por la distancia del color al eje central.

En lo que al color como variable visual respecta, cada una de estas componentes de un color son a su vez variables visuales, y como tales pueden emplearse para simbolizar los distintos elementos de un mapa. En la práctica, el tono y el valor son utilizadas frecuentemente, pero la saturación tiene una utilidad muy limitada, por lo que es infrecuente su uso. En lo sucesivo, por tanto, trataremos el color no como una única variable visual sino como dos distintas: valor y tono.

27.2.4. Tamaño

El tamaño se refiere a la dimensión del símbolo. Para el caso de símbolos puntuales, puede aplicarse sin más que hacer más grande o pequeño el símbolo en sí. En el caso de líneas, modificar su grosor constituye la forma de aplicar la variable tamaño. Al igual que sucedía con la forma, en las superficies va a implicar la modificación de estas, por lo que se emplea únicamente en los cartogramas. Otra forma de aplicar el tamaño a los símbolos superficiales es hacerlo sobre la textura con la que estos se rellenan, usando un único patrón con diferentes tamaños en sus tramas (Figura 27.3).

El tamaño condiciona la percepción de otras variables visuales, especialmente cuando se trata de tamaños pequeños. Un punto muy pequeño o una línea demasiado fina no van a permitir la aplicación de, por ejemplo, el tono o el valor, o al menos no del mismo modo que con un tamaño mayor, ya que la percepción de estas variables será más difícil.

27.2.5. Textura

La textura hace referencia al relleno de un símbolo mediante algún patrón. Empleando patrones distintos se produce una diferenciación en los símbolos correspondientes.

En el caso de los símbolos puntuales, la textura requiere que estos tengan un tamaño suficiente para que pueda apreciarse el patrón que constituye cada una de las texturas. Este tamaño mínimo requerido es mayor que en el caso de emplear el tono o el valor.

En el caso de líneas, entendemos como textura el uso de guiones y espacios en blanco que dan lugar a un patrón de discontinuidad, como se muestra en la figura 27.4. No obstante, esta discontinuidad es una desventaja a la hora de representar un elemento lineal, ya que implica que una parte de él no va a representarse. Dependiendo del significado de aquello que representemos, el uso de texturas en elementos lineales puede no ser lo más recomendable a la hora de crear un mapa. Puede emplearse otro tipo de texturas para formar líneas, «rellenando» estas si tienen

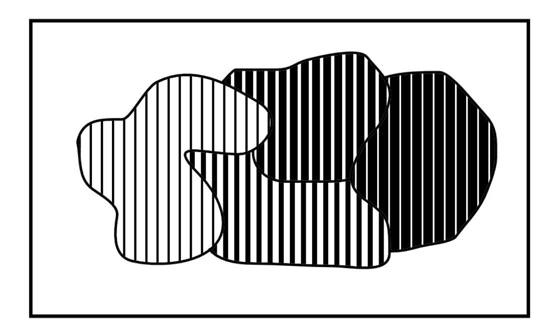


Figura 27.3: Uso del tamaño en símbolos de superficie mediante texturas.

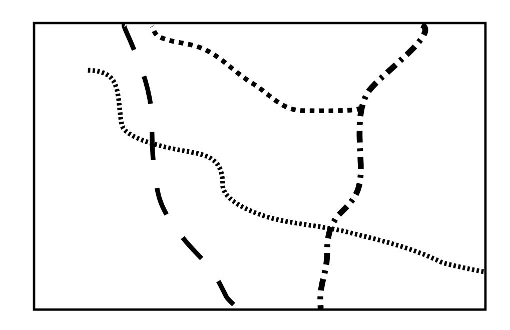


Figura 27.4: Aplicación de la variable visual textura a los símbolos lineales.

un grosor considerable, pero su uso no se recomienda.

Las texturas se aprovechan plenamente sobre los símbolos de superficie, ya que la mayor dimensión de estos permite una percepción completa y una interpretación mucho más sencilla, al igual que ocurría en el caso del valor.

27.2.6. Orientación

La última variable visual es la orientación. Se aplica sobre los símbolos puntuales, siempre que estos no presenten simetrías que impidan percibir correctamente la orientación. Por ejemplo, para el caso del círculo, resulta obvio que no tiene sentido aplicar la orientación como variable visual. Los símbolos compuestos por formas geométricas son adecuados para emplear la orientación, mientras que los símbolos pictóricos no responden de igual forma y producen en la representación sensación de desequilibrio. Se recomienda, por tanto, emplear esta variable únicamente con los primeros.

Puede aplicarse también sobre los símbolos de superficie a través de la textura, variando la orientación de esta. Sobre las líneas, su aplicación no es posible. Puede emplearse en caso de líneas con textura, pero esto requiere un ancho excesivo para una correcta percepción.

27.3. Las propiedades de las variables visuales

Las variables que acabamos de ver son ahora las herramientas que emplearemos para simbolizar la información geográfica, y sabemos ya cómo aplicarlas. Lo que no hemos visto aún es qué capacidades tienen y qué podemos simbolizar mediante ellas, y este es realmente el aspecto clave sobre el que deberemos decidir posteriormente cuando nos dispongamos a crear un mapa, para así seleccionar la variable visual más adecuada en función de aquello que queramos representar.

Se distinguen cuatro propiedades básicas que una variable visual puede presentar:

- **Asociativa**. Una variable visual presenta la propiedad asociativa si al ser aplicada no aumenta ni disminuye la visibilidad de un elemento. Es decir, cuando en función de esa variable visual no puede asignársele más o menos importancia a este.
- **Selectiva**. La propiedad selectiva la presentan aquellas variables visuales que, al ser aplicadas, generan distintas categorías de símbolos.
- **Ordenada**. Cuando una variable visual puede emplearse para representar un orden, se dice que presenta la propiedad ordenada.

- **Cuantitativa**. Cuando, además del orden, una variable puede mostrar cantidades o proporciones, entonces se dice que posee la propiedad cuantitativa.

El orden en que se han presentado estas propiedades no es casual, ya que están ordenadas dando lugar a lo que Bertin denomina *niveles de organización*. La propiedad asociativa se sitúa en el nivel más bajo, mientras que la cuantitativa ocupa el más alto. El nivel de organización de las variables visuales tiene importancia a la hora de combinar varias de ellas en un símbolo, como veremos más adelante. Asimismo, y como detallaremos en el capítulo siguiente, el nivel de organización define qué tipo de información podemos transmitir con una variable visual.

Para ver más exactamente el significado de estas propiedades, estudiemos con detalle la figura 27.5, que muestra diferentes representaciones de un conjunto de símbolos (en este caso, símbolos puntuales) en los que en cada caso se ha utilizado únicamente una variable visual.

Comenzando con la propiedad asociativa, vemos que a excepción del tamaño y el valor, las demás variables visuales no hacen que los elementos presenten una preponderancia en la imagen. No existen una orientación que podamos definir como más importante, ni tampoco un color. Lo mismo sucede con la textura, la forma y la posición. Podemos emplear una u otra forma, o una u otra textura, y con ello no conseguiremos llamar más la atención sobre un elemento en cuestión.

Con el tamaño, sin embargo, resulta claro que mayor tamaño implica un papel destacado dentro de la información que transmite el mapa. De igual modo, un mayor valor (un color más oscuro) da sensación de mayor definición, y centra la atención de observador sobre el elemento de un modo muy superior a como lo hace un valor bajo.

Respecto a la propiedad selectiva, diremos que una variable visual la presenta si de un vistazo podemos rápidamente seleccionar los elementos que pertenecen a un determinado grupo, identificados estos mediante dicha variable visual. El caso más claro de propiedad selectiva lo presenta el tono. Podemos rápidamente quedarnos solo con los elementos amarillos o con los rojos. Aunque no de un modo tan claro, todas las restantes variables presentan igualmente esta propiedad, a excepción de la forma. La forma no permite que los elementos se agrupen de modo espontáneo en familias, y su validez en este sentido está muy ligada a la complejidad de dicha forma.

La propiedad ordenada la presentan aquellas variables que permiten establecer un orden. Tan solo posición, textura, tamaño y valor la presentan, mientras que las demás carecen de ella. Por ejemplo, en la imagen correspondiente a la variable visual tono no podemos decir cuáles de los elementos situaríamos al principio y cuáles al final de una escala dada definida por esos tonos. Con el valor, sin embargo, sí que podemos, ya que esta escala iría de los tonos más claros a los más oscuros, y visualmente podemos sin dificultad distinguir los distintos niveles y ordenarlos.

Por último, la propiedad cuantitativa la presentan aquellas variables visuales que permiten estimar proporciones o cantidades de forma visual. Esta propiedad es exclusiva del tamaño y de la posición, mientras que las demás no la presentan. Podemos visualmente estimar una distancia en comparación con otra y decir que es, por ejemplo, el doble de esta. También podemos ver que los círculos grandes en la figura correspondiente son aproximadamente el doble que los pequeños.

El valor, que ya sabemos que presenta la propiedad ordenada, podría pensarse que también presenta la propiedad cuantitativa, pero no sucede así. Es difícil e impreciso afirmar que un color

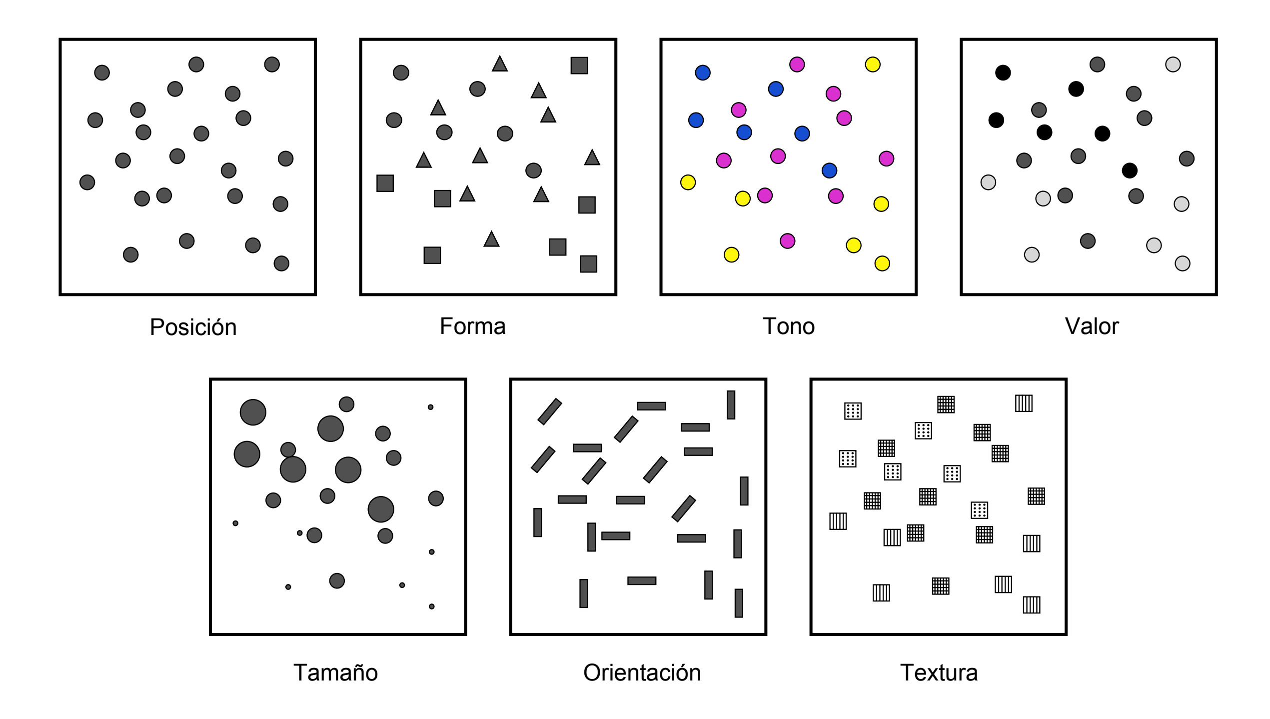


Figura 27.5: Representación de un conjunto de símbolos aplicando de forma individual las distintas variables visuales.

es el doble de oscuro que otro, y lo más que podemos hacer es situarlo entre dos valores distintos (de ahí que posea la propiedad ordenada), pero no deducir una cifra que exprese una cantidad o proporción. Las restantes variables visuales resulta claro que no poseen esta propiedad.

En el cuadro 27.3 se muestra un resumen de todo lo anterior.

Aunque las ideas de Bertin conforman una sólida base teórica de reconocido valor, lo cierto es que debe permitirse cierta laxitud en la aplicación de estas, y no considerar que existe una dicotomía estricta en el caso de las propiedades antes presentadas. Hay muchos factores y circunstancias que pueden alterar la forma en que estas propiedades se presentan, y alterar la intensidad con que aparecen en unas u otras variables visuales.

Por ejemplo, aunque el tono no presenta, según la propuesta original de Bertin, la propiedad ordenada, sí que puede emplearse para representar un orden en determinadas circunstancias. Si estamos simbolizando unos valores de temperatura, podemos establecer una transición de colores entre el rojo y el azul, que serán fácilmente identificados y ordenados por el observador del mapa, ya que el primero de estos colores se asocia habitualmente al calor y el segundo al frío. En este contexto particular, el tono sí presenta la propiedad ordenada.

En los capítulos 15 o 14 verás muchos ejemplos de representaciones en que se usan gradaciones de tono para simbolizar variables de tipo cuantitativo, ya sean razones o proporciones. Estas guardan, no obstante, cierta lógica, de tal modo que puede entenderse adecuadamente su

	Posición	Tamaño	Forma	Valor	Tono	Textura	Orientación
Asociativa	♢	-	♢	-	♢	♢	♢
Selectiva	♢	♢	-	♢	♢	♢	♢
Ordenada	♢	♢	-	♢	-	-	-
Cuantitativa	♢	♢	-	-	-	-	-

Cuadro 27.1: Cuadro resumen con las propiedades de las variables visuales.

significado. Como veremos en el próximo capítulo, esto también tiene relación con el tipo de mapa, de tal modo que ciertos tipos de mapas permiten por sus propias características el uso del tono para este tipo de variables.

Junto a lo anterior, algunos autores (véase [329]) expanden el número de variables visuales y se han desarrollado revisiones a las propiedades enunciadas por Bertin basadas en estudios prácticos, que demuestran cómo pueden existir variaciones sobre la relación entre estas y las distintas variables visuales (por ejemplo, [330]).

27.4. Uso combinado de las variables visuales

Para explicar cada una de las variables visuales, hemos visto diversos ejemplos en los que utilizábamos cada una de ellas por separado y de forma única. Sin embargo, las variables visuales pueden combinarse y, si se hace de la manera correcta, esto reforzará la capacidad que estas tienen para transmitir una información dada. La imagen 27.6 muestra algunos ejemplos de combinación de variables visuales que nos servirán para detallar la forma adecuada de usar varias de ellas simultáneamente.

El primero de los ejemplos propuestos muestra el uso combinado de las variables tamaño y forma para símbolos puntuales. Supongamos que estos símbolos representan la profundidad del suelo medida en determinados emplazamientos, estando relacionado un mayor tamaño del símbolo con una profundidad mayor. Asimismo, se ha asociado un símbolo triangular a los valores más bajos, un símbolo circular a los intermedios y uno cuadrado a los más altos.

Aunque se emplean dos variables visuales distintas, el resultado no es, sin embargo, mejor que en caso de emplear uno solo de ellos (en este caso, debería emplearse el tamaño, ya que la forma no presenta la propiedad cuantitativa necesaria para representar cantidades). Lejos de producirse una sinergia entre el efecto de ambas variables, el resultado es similar al uso exclusivo del tamaño en cuanto a su capacidad de transmitir la información, o incluso peor, ya que la forma puede dificultar la estimación visual del tamaño, al ser más complicado comparar la dimensión de objetos de distinta forma.

Pese a que no es clara la ventaja de aplicar conjuntamente las variables forma y tamaño, esta puede emplearse para representar cantidades, por lo que podemos decir que mantiene la propiedad cuantitativa que posee el tamaño. En general, al combinar dos variables visuales el resultado presentara las propiedades de aquella que tenga un mayor nivel organizativo. Puesto que la propiedad cuantitativa representa el nivel orga-

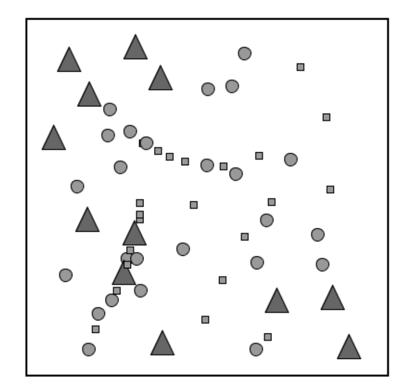
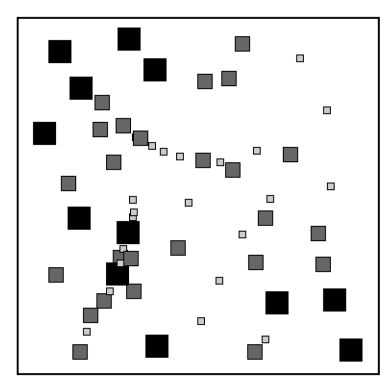
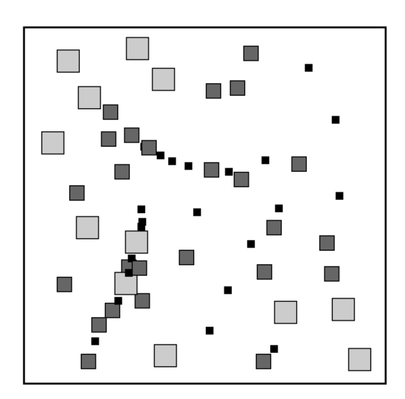




Figura 27.6: Combinación de variables visuales.

nizativo superior, en este caso se mantiene en la combinación.

Aún así, hay mejores formas de combinar las variables visuales para que esta combinación enfatice en mayor grado la información que se pretende transmitir, como por ejemplo la mostrada en el segundo ejemplo. Este ejemplo combina el tamaño y el valor, variables ambas que no poseen la propiedad asociativa. Es decir, poseen su complementaria, que podríamos denominar *disociativa*, y que, recordemos, es la propiedad que, al aplicarse sobre un símbolo, hace que este gane importancia visual. El resultado presenta un carácter todavía más disociativo, en cuanto que los símbolos que representan una cantidad elevada, al ser no solo grandes, sino estar pintados en color oscuro, llaman aún más nuestra atención que si empleáramos una única de las variables visuales utilizadas.

Como regla en este sentido, podemos decir que, cuando se combinan variables visuales que poseen una determinada propiedad, en el resultado esta propiedad queda reforzada con respecto a las variables individuales.

El tercer ejemplo nos muestra que combinar variables visuales con una misma propiedad no garantiza necesariamente que se vaya a producir una sinergia entre ellas, sino que, por el contrario, pueden anularse. Las variables empleadas en este caso son las mismas, valor y tamaño, pero se ha asociado el color claro a los valores mayores y el oscuro a los menores, de tal modo que los símbolos de mayor tamaño son más claros que los pequeños. Esto atenúa el efecto disociativo del tamaño, de forma que la representación es más difícil de interpretar y su información no se transmite de modo tan inmediato y directo.

En resumen, podemos sintetizar lo anterior diciendo que, a la hora de combinar variables visuales, deben tenerse en cuenta las propiedades de estas del mismo modo que cuando se emplean de forma individual. Las propiedades a reforzar serán aquellas que convengan más al tipo de información representado, y deben presentarlas todas las variables a combinar para que el efecto conjunto sea más acusado.

27.5. La percepción visual

La percepción engloba toda la serie de procesos que convierten un fenómeno físico en una información acerca de nuestro entorno, a través de la estimulación de unos órganos perceptivos.

La percepción tiene una fase física, una fisiológica (la estimulación en sí) y una psicológica (la interpretación del estímulo). En el caso de la percepción visual, este fenómeno físico es de tipo energético (la luz), y los órganos correspondientes son los ojos.

El estudio de la percepción es un fenómeno complejo que no entraremos a detallar, pero en el que resulta de interés profundizar para conocer algo más acerca de cómo la información que plasmamos en un mapa (que es un elemento visual) acaba convertida en una información en la mente del observador de ese mapa. Entender este proceso, al menos someramente, nos permitirá mejorar la eficacia de la percepción, de forma que tengamos una mayor garantía de que la información que transmitimos sea recibida e interpretada correctamente.

Dos son los aspectos que detallaremos en esta sección: las constancias perceptivas y las ayudas a la percepción. En otras palabras, hasta qué punto podemos modificar los elementos visuales o su entorno sin que dejen de transmitir su información y sean confundidos sus características, y cómo podemos facilitar que se perciban exactamente como pretendemos.

27.5.1. Las constancias y contrastes perceptivos

Entendemos por constancias perceptivas a las propiedades de los objetos cuya percepción no varía aunque se produzcan modificaciones. Podemos ver algunos ejemplos para algunas de las variables visuales que conocemos.

Dado un objeto redondo tal como una rueda, si lo miramos en una dirección perpendicular aparecerá efectivamente como una forma circular perfecta. Sin embargo, si la miramos desde otro ángulo, veremos una forma elíptica, pero ello no nos lleva a pensar que la rueda en sí no sea ya redonda. Nuestra percepción de esa rueda es la misma, y podemos apreciar de igual modo su tamaño o su forma. Alterar el ángulo de visión no altera el objeto y la percepción que tenemos de él.

Del mismo modo, un elemento pintado de un color claro se identifica como tal aunque la luz sea tenue, y un elemento oscuro lo seguimos percibiendo como oscuro aunque estemos en unas condiciones de iluminación fuerte. Nuestro cerebro es capaz de interpretar simultáneamente el objeto y el contexto, y de este modo extraer las características de ese objeto, que no varían.

Estos dos ejemplos muestran la constancia perceptiva de la forma y el valor, y podemos buscar otros similares para otras variables visuales.

No todas las variables visuales tienen una constancia perceptiva como la anterior. Todos conocemos múltiples ejemplos de ilusiones ópticas en las que algo no parece lo que realmente es, y esa percepción errónea viene normalmente motivada por las condiciones en las que percibimos el objeto, por ejemplo debido al entorno particular en el que este se encuentra junto a otros objetos. La figura 27.7 muestra un ejemplo clásico de ilusión óptica, conocida como *ilusión de Zollner*. Las líneas largas diagonales son paralelas, pero no aparentan serlo, debido al efecto causado por las líneas más cortas. En este caso, no existe una constancia perceptiva de la variable visual orientación.

Cuando la percepción de un elemento cambia aunque el estimulo no lo haga, en lugar de una constancia perceptiva hablamos de un *contraste perceptivo*. Los contrastes perceptivos son importantes, ya que pueden inducir una interpretación errónea de la información que pretendemos transmitir, al producirse una percepción equivocada.

Las siguientes son algunas de las ideas más importantes a tener en cuenta a este respecto a la hora de crear un mapa:

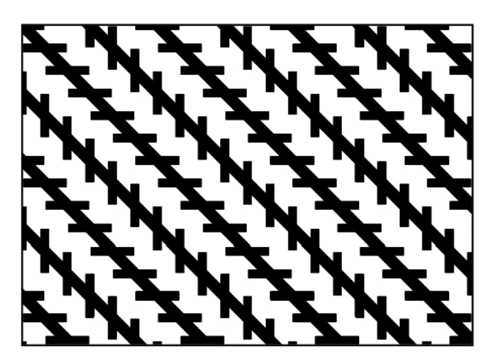


Figura 27.7: Ilusión de Zollner que demuestra el contraste perceptivo de la orientación.

- El tamaño es la variable visual que más afectada se ve, y el tamaño aparente de un objeto puede variar notablemente si se encuentra rodeado de otros de un tamaño distinto. La figura 27.8 muestra un ejemplo de esto. A la hora de emplear simbología de elementos puntuales en un mapa (por ejemplo, en un mapa de símbolos graduados, como veremos en el apartado 28.6.1), esto debe tenerse en cuenta, ya que pueden presentarse situaciones como la de la figura.
- El valor se ve igualmente alterado al situar alrededor elementos de distinto valor. Si el número de distintos valores es pequeño, es más difícil que aparezca este contraste perceptivo. A medida que se aumenta el número de estos, es más probable que aparezca en mayor o menor medida.
- El tono se ve alterado por la presencia de otros tonos distintos. En un mapa, veremos este efecto al enfrentar el color de un elemento sobre el color del fondo. Por ejemplo, si una línea que representa a una carretera cruza una serie de polígonos de distinto tono, puede parecer que el tono de la línea varia aunque en realidad sea constante.
- Tonos complementarios puestos juntos pueden crear sensación de vibración en la frontera que los separa.

27.5.2. Ayudas a la percepción

Con lo que hemos visto anteriormente, queda claro que podemos alterar la forma en que se perciben las variables visuales que caracterizan a un elemento visual. Podemos usar este hecho para nuestro beneficio, de tal modo que el diseño de un mapa incorpore elementos que hagan más patente la información que este contiene, facilitando la correcta percepción del mapa en su conjunto.

Un factor clave en este sentido es la adecuado separación entre el fondo y la figura. Aquello que queremos que resulte visible con carácter principal (en el caso de un mapa, sus distintos elementos) debe separarse de aquello que constituye el fondo de la imagen, y debe atraer la atención del observador de manera prioritaria. En caso de no ser así, puede resultar difícil «descubrir» la información que el mapa transmite, al quedar esta al mismo nivel que la de otros elementos de menor importancia. El ejemplo clásico de la figura 27.9 ilustra este hecho. Puesto que no existe una diferenciación clara entre el fondo y la figura, no es obvio saber si la imagen pretende representar una copa o dos caras.

En un mapa, y como veremos en el próximo capítulo, encontramos dos tipos de cartografía: una con carácter de base que define un contexto

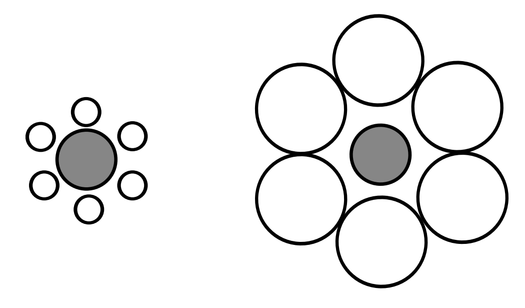


Figura 27.8: Contraste perceptivo del tamaño. Ambos circulos grises tienen el mismo tamaño, pero el de la izquierda aparenta ser mayor.

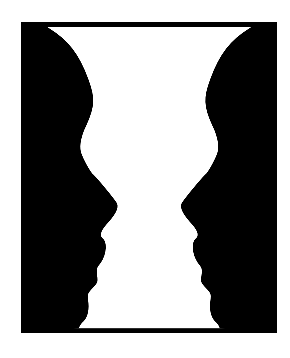
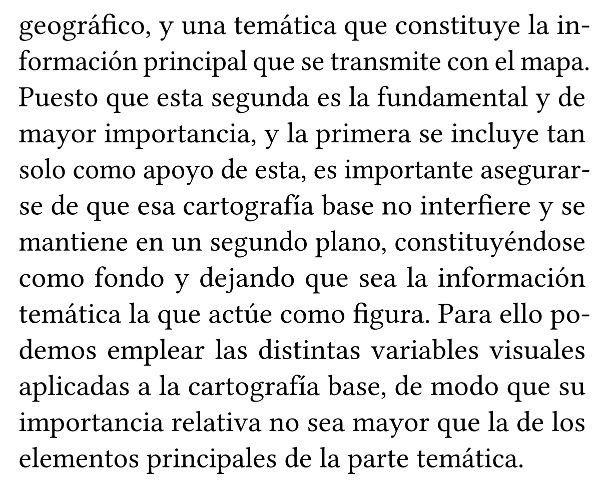


Figura 27.9: Sin un adecuado contraste entre fondo y figura la imagen presenta ambigüedad.



geográfico, y una temática que constituye la información principal que se transmite con el mapa. Puesto que esta segunda es la fundamental y de mayor importancia, y la primera se incluye tan solo como apoyo de esta, es importante asegurarse de que esa cartografía base no interfiere y se mantiene en un segundo plano, constituyéndose como fondo y dejando que sea la información temática la que actúe como figura. Para ello podemos emplear las distintas variables visuales aplicadas a la cartografía base, de modo que su importancia relativa no sea mayor que la de los elementos principales de la parte temática.

Otro aspecto a considerar es la adecuada jerarquización entre los elementos del mapa. La división entre fondo y figura ya constituye en sí una jerarquización, pero no es suficiente si conviven varios tipos de elementos en el mapa. Dentro de la parte temática es necesario estructurar estos visualmente para que quede clara su importancia y se vea sin dificultad que existe una división entre ellos.

Esta jerarquía debe aportar una «profundidad» a la información, de forma que existan niveles en esta y se perciba que algunos elementos están por encima de otros. Como veremos en el capítulo 29, la forma de ordenar las distintas capas en un SIG ya establece un orden, aunque este no es en sí suficiente, y deben utilizarse las variables visuales para enfatizar o no unas o otras capas y la información que contienen.

Algunas técnicas básicas para esto son las que permiten que exista algún factor diferencial en la información más relevante. Si las propiedades de los elementos destacados difieren notablemente de las del fondo, esto centra la atención sobre ellas y garantiza que no se confundan con este. Emplear unas características más homogéneas para el fondo permite que la diferenciación de la figura sea más patente. En otras palabras, el contraste, aplicado este a todas las variables visuales, es una de las claves para lograr una adecuada transmisión de la información al emplear una representación visual.

El contraste se aplica no solo a las variable visuales, sino en general a las características de la representación. Por ejemplo, el nivel de detalle es una propiedad susceptible de ser utilizada para enfatizar algo. Así, y en el caso particular del documento cartográfico, el lector de un mapa espera que el detalle sea mayor en la cartografía temática que en la de base, ya que esta última es simplemente un elemento complementario de ayuda. Un mayor detalle sobre ciertos elementos llamará más la atención en contraste con un fondo menos detallado, y esto puede utilizarse

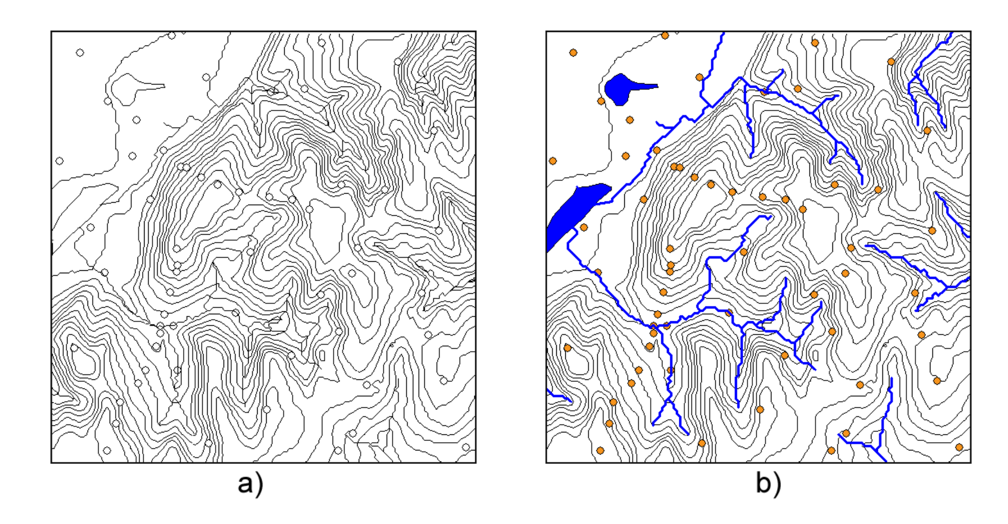


Figura 27.10: Mapa con jerarquía incorrecta (a) y mapa adecuadamente jerarquizado (b).

para enfocar la atención sobre lo más relevante. Ofrecer menos detalle en la cartografía de base no es un inconveniente si esto ayuda a un mejor entendimiento de los elementos principales del mapa.

Como ejemplo de lo anterior, la figura 27.10 muestra un ejemplo de cómo una correcta jerarquización es fundamental para crear mapas de calidad.

Por último, un aspecto clave para la claridad de un mapa es el relativo al poder separador. Este define la capacidad de un individuo para distinguir objetos muy pequeños y separar objetos cercanos. Además de depender del propio individuo, está condicionado por una serie de factores.

Se admite en líneas generales que el límite de separación entre dos objetos para el ojo humano es de 0,2mm. Si existe una distancia menor entre ellos, en condiciones normales no será posible distinguir uno de otro.

Existe también un límite para poder reconocer objetos aislados, aunque este depende del tipo de objeto. Los siguientes son algunos de los aplicados usualmente:

- 0,2mm de diametro para el caso de un punto.
- 0,5mm de grosor para el caso de una línea negra.
- 0,4mm de lado para el caso de un cuadrado negro.
- 0,6mm de lado para un cuadrado sin relleno.

Asimismo, hay un umbral de diferenciación, que define el tamaño mínimo de dos objetos para que puedan ser percibidos como distintos. Este umbral también depende de las caracteristicas de los objetos, como por ejemplo la forma (si las formas son muy distintas será más fácil distinguirlos que si son muy similares).

El poder separador no depende únicamente de variables de tipo espacial, sino que también está en relación con otras variables visuales. Por ejemplo, una línea negra sobre fondo blanco puede distinguirse aunque sea fina, pero en caso de ser amarilla sobre ese mismo fondo, será necesario un grosor mayor.

Como parece lógico, estos conceptos deben usarse para no incorporar a un mapa elementos que estén más allá del umbral de separación del lector del mapa, ya que en tal caso no podrá extraer la información de este.

27.6. Resumen

Para transmitir correctamente cualquier tipo de información mediante el lenguaje visual, es necesario conocer sus elementos y saber emplearlos de modo adecuado. La semiología gráfica se encarga del estudio de los símbolos del lenguaje visual, y en este capítulo hemos visto algunas de sus ideas principales.

De especial relevancia resultan las denominadas *variables visuales*, las cuales empleamos para la caracterización de símbolos. Existen seis variables visuales: posición, forma, tamaño, color, textura y orientación. El color a su vez se puede dividir en tres: tono, valor y saturación. De estas tres, solo las dos primeras, tono y valor, tienen aplicación práctica en el ámbito cartográfico.

Las variables visuales presentan distintas propiedades, que definen a su vez los *niveles de organización*. De menor a mayor organización, estas propiedades son las siguientes: asociativa, selectiva, ordenada y cuantitativa. Las propiedades de una variable visual condicionan el tipo de información que puede transmitirse haciendo uso de ella. Cuando se combinan varias variables visuales que poseen una misma propiedad, esta propiedad se presenta con mayor fuerza en el resultado.

Podemos ayudar a que la percepción de la información que transmitimos con un elemento visual sea mejor, atendiendo a aspectos como el contraste entre el fondo y la figura, así como estableciendo una correcta jerarquización entre los distintos elementos. Igualmente, debemos prestar atención a los contrastes perceptivos, para evitar que estos aparezcan y se produzca una percepción incorrecta.

El mapa y la comunicación cartográfica

28.1. Introducción

Los mapas se han empleados desde la antigüedad para recoger la información geográfica y transmitirla. Como ya dijimos en el capítulo anterior, podemos entender un mapa como un medio de comunicación visual que constituye un lenguaje con un objetivo particular: la descripción de relaciones espaciales. Una mapa es, pues, una abstracción simbólica de algún fenómeno real, lo cual significa que presenta un cierto grado de simplificación y generalización.

El diseño, producción y uso de un mapa como forma de comunicación conforma lo que se conoce como *proceso cartográfico*. Más concretamente, el proceso cartográfico conlleva cuatro etapas o subprocesos, a saber:

- Recogida de los datos.
- Manipulación y generalización de los datos para diseñar y construir el mapa.
- Visualización del mapa.
- Interpretación de la información.

La labor del cartógrafo se centra en el segundo de estos puntos, mientras que el usuario del mapa lleva a cabo los dos últimos. Será en esa construcción de los mapas en lo que nos fijemos a lo largo de este capítulo, para conocer los conceptos y reglas que rigen la comunicación cartográfica a través del uso de mapas.

El lenguaje visual que estudiábamos en el capítulo 27 se convierte ahora en un lenguaje cartográfico al adaptarlo al caso particular de la creación de mapas, y estas reglas (equivalentes a la gramática y la sintaxis de un lenguaje hablado) son imprescindibles para poder crear cartografía que facilite las citadas labores del usuario posterior de esta. Este conjunto de ideas relativas a la producción de mapas dan forma a lo que conocemos como *diseño cartográfico*.

El diseño cartográfico implica la toma de decisiones por parte del cartógrafo. Algunas de estas decisiones pueden ser la cantidad de simplificación que debe realizarse o los símbolos que han de emplearse para plasmar la información a transmitir. Las ideas desarrolladas en los próximos apartados conforman una base de conocimientos que facilita la toma de decisiones correctas en este sentido.

28.2. El propósito del mapa

Como elemento de comunicación, un mapa tiene siempre un propósito. De la misma forma que al hablar pretendemos transmitir algo y para ello usamos el lenguaje como herramienta, en el caso de crear un mapa empleamos el lenguaje gráfico para transmitir una determinada información geográfica. También de igual modo que en el caso de la comunicación verbal, y en el de cualquier otra forma de comunicación, existe un receptor de nuestro mensaje. Es decir, un usua-

rio (o varios) de ese mapa, que serán quienes lo interpreten y aprovechen.

Esto que parece obvio es un hecho que se ignora muchas veces a la hora de elaborar un mapa, y con ello se pierde gran parte de la capacidad del mapa como elemento de comunicación. Aplicar los conceptos de visualización correctamente, así como aquellos que veremos en este capítulo relativos a la simbolización, no garantiza que el mapa que generemos sea útil, del mismo modo que aplicar adecuadamente la gramática del chino para elaborar una frase no sirve de nada si nuestro interlocutor solo habla castellano. No será capaz de interpretar nuestro mensaje por muy correcto que este sea. Resulta incluso mejor elaborar un mensaje con errores gramaticales en castellano, ya que al hacerlo así estamos teniendo en cuenta las circunstancias en que se produce la comunicación.

Al crear un mapa nunca debemos olvidar quién y para qué va a usar ese mapa, y en función de ello elegir los elementos correctos y la forma de presentar la información más acorde con esos destinatarios y sus objetivos particulares. Sólo entonces es cuando aplicaremos los conceptos del diseño cartográfico para que el mensaje que elaboramos sea el mejor posible.

La figura 28.1 muestra un ejemplo claro de lo anterior a través de dos mapas con predicciones meteorológicas, proporcionados por la Agencia Estatal de Meteorología de España. El primero es un mapa de probabilidad de precipitación, mostrada esta mediante isolíneas. El segundo es un clásico mapa del tiempo (conocido como *mapa significativo*) en el que sobre el mismo territorio se sitúan símbolos indicando el tiempo previsto (soleado, chubascos, lluvias, tormentas, etc.).

Ambos mapas son correctos desde el punto de vista de la labor cartográfica y se han creado a partir de una misma información, pero la forma de mostrar esta es bien distinta. Para un uso científico, este último mapa resulta claramente insuficiente, mientras que el primero es adecuado. Sin embargo, si la audiencia es no especializada, tal como los lectores de un periódico que deseen saber si mañana podrán o no salir al campo a disfrutar de un día soleado, el segundo mapa es mucho mejor, ya que el primero, aunque también proporciona esa información e incluso lo hace con más detalle, puede resultar excesivamente complejo y difícil de entender si no se tienen ciertos conocimientos. Es decir, el usuario es en última instancia, y por encima del propio diseño cartográfico, quien hace que el mapa sea o no un elemento útil.

Entre los elementos fundamentales que se han de elegir en función del propósito del mapa se encuentran los correspondientes a la base matemática del mapa: escala y proyección. La escala condicionará el tipo de estudios que será posible llevar a cabo con el mapa, y establecerá el nivel de detalle que se desea comunicar a través de este (siempre, obviamente, dentro de los limites de la escala a la que se hayan recogido los datos). Por su parte, la proyección debe considerarse en función de sus propiedades. Como ya vimos en el apartado 3.3.2, toda proyección implica algún tipo de distorsión. Existen así proyecciones que mantienen las áreas, las distancias o los ángulos. Según qué trabajo se espere con el mapa, será más indicado hacer uso de una u otra de ellas, ya que no es lo mismo un mapa catastral que una carta de navegación, y la elección de una proyección inadecuada puede convertir un mapa en una herramienta inútil para la tarea que se pretende realizar.

El otro aspecto importante a considerar es la forma en que transmitimos la información a través del mapa, es decir, el tipo de mapa, como hemos visto en el ejemplo propuesto. Dentro de este capítulo estudiaremos los tipos de mapas más

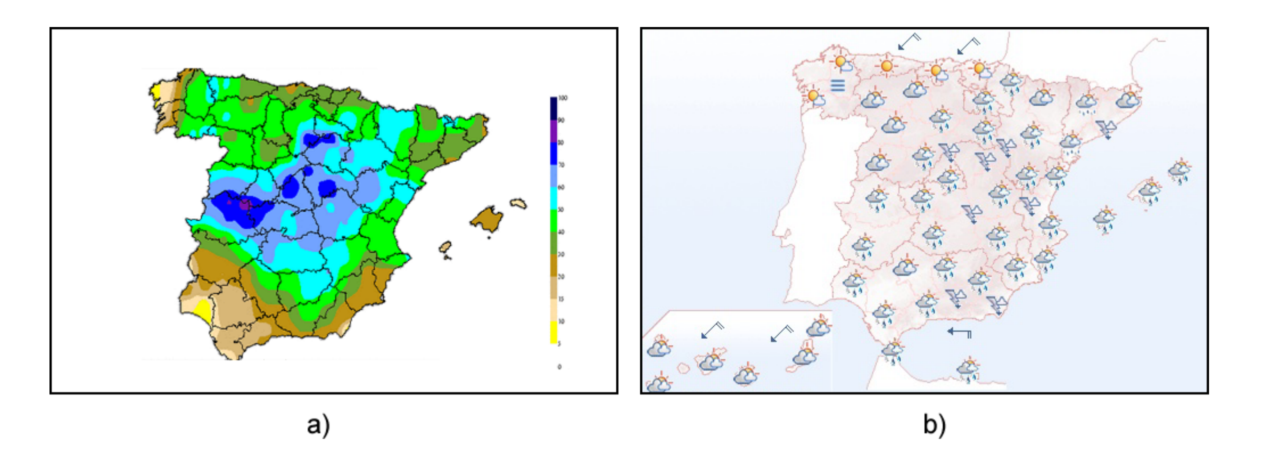


Figura 28.1: Dos formas distintas de mostrar una información a través de un mapa. En función del propósito de este y el publico al que va dirigido, cada una de ellas podrá ser adecuada o no. (Imágenes cortesía de AEMET)

habituales y las características que los definen, así como la forma de crearlos correctamente.

28.3. Cartografía temática y cartografía base

Existen muchos tipos de mapas y muchas formas de clasificarlos. Una clasificación especialmente relevante es la que divide a estos en dos grupos cartográficos principales en función del tipo de información que aporten: *cartografía base*, también denominada *fundamental* o *topográfica*, y *cartografía temática.*

La cartografía base representa el tipo de mapa que originalmente era el objeto principal de la cartografía, cuando lo primordial era recoger con precisión *qué* había sobre la Tierra, documentando a través del documento cartográfico las características físicas de esta. Este tipo de cartografía requiere de medidas precisas y se basa fundamentalmente en el trabajo de la topografía para obtener la información necesaria que posteriormente se plasma sobre el mapa.

La cartografía base tiene carácter general, y ello explica que inicialmente fuera el único tipo de mapa de interés para el cartógrafo, ya que existía una indudable necesidad de ese tipo de información de referencia acerca del entorno físico. Una vez que se ha desarrollado una colección suficiente de mapas topográficos y se conoce bien la Tierra a través de ellos, los cartógrafos comienzan a recoger en otro tipo de mapas otras variables espaciales también susceptibles de ser representadas de ese modo. Esto tiene lugar alrededor del siglo XVIII, y aparece entonces la cartografía temática.

La cartografía temática se centra en la representación de un tema concreto (una variable espacial dada), pudiendo ser esta de cualquier índole: física, social, política, cultural, etc. Se excluyen de la lista de esos temas posibles a los puramente topográficos, que constituyen el objeto de la cartografía base.

La cartografía temática se apoya en la cartografía base, ya que esta se incluye también en los mapas temáticos para facilitar la comprensión del comportamiento espacial de la variable repre-

sentada y ubicar esta en un contexto geográfico dentro del propio mapa. Un mapa temático se compone, así pues, de dos partes bien diferenciadas:

- Una capa específica con la información temática. Contiene la información principal del mapa, que representa la variable espacial sobre la que se construye este.
- Un mapa base. El mapa base provee una localización geográfica a la que se referencia la información temática. Debe contener los elementos propios de la cartografía base, aunque siempre ha de tenerse en cuenta que estos han de coexistir con los correspondientes a la parte temática. Por ello, frecuentemente es necesario incluir en este mapa base menos detalle que si se diseñara para ser un mapa independiente, limitándose a los elementos necesarios que definan un contexto geográfico básico. La labor de este mapa base no es ser utilizado como tal como si se tratara de cartografía base aislada, sino ayudar a los elementos de la componente temática a transmitir mejor la información que contienen.

  Aunque en ocasiones puede utilizarse un mapa topográfico estándar como mapa base, habitualmente este contiene demasiada información e interfiere con la capa temática, siendo más adecuado crear el mapa base a partir de elementos individuales. Algunos de los más importantes son el *canevás* (rejilla de coordenadas, especialmente necesaria a escalas pequeñas), la red fluvial, el relieve, la vías de comunicación, las poblaciones y los nombres geográficos. Todos ellos son buenos elementos de referencia para permitir situar en base a ellos cualquier tipo de información temática.

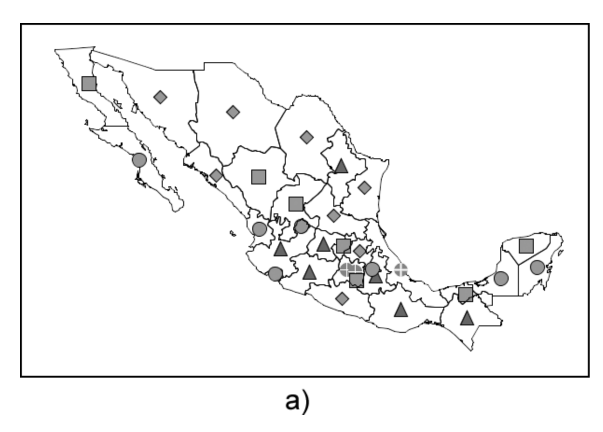
La mayoría de las ideas de este y el próximo capítulo se aplican fundamentalmente a la cartografía temática, siendo esta además la que con mayor frecuencia se genera mediante el uso de un SIG. Una buena parte de lo visto en relación con las variables visuales y sus propiedades tiene mayor relevancia a la hora de tratar con cartografía temática, ya que esos conceptos se aplican a la representación de variables y fenómenos de tipo cuantitativo, y es la cartografía temática la que trabaja con ellos.

En la cartografía topográfica, los elementos geométricos que representamos son en sí la información que pretendemos comunicar con el mapa, mientras que en la cartografía temática esa geometría es solo parte de la información, siendo la otra parte la que se transmite a través del uso de variables visuales como, por ejemplo, el color. De otro modo, la cartografía topográfica representa «cosas» que encontramos en el terreno (un accidente geográfico, el curso de un río, el perfil de una costa), mientras que la cartografía temática se centra más en la representación de valores y atributos. La línea que representa una carretera en un mapa existe realmente en el terreno, mientras que la que representa una curva de nivel no existe físicamente. Podemos decir también que en lugar de en el *qué*, la cartografía temática se centra en el *cómo*.

Según el tipo de información que contenga, la cartografía temática se divide en cuantitativa y cualitativa. Como veremos a continuación, el tipo de información tiene gran repercusión a la hora de generar un mapa, ya que condiciona los elementos que podemos usar para simbolizar dicha información.

28.4. Los tipos de información y su representación

Como vimos en el apartado 4.3, la componente temática de la información geográfica puede ser de tipo numérico o alfanumérico, y la primera se divide en los tipos nominal, ordinal, intervalos y razones. Nominal y alfanumérico representan información cualitativa, mientras que los restan-



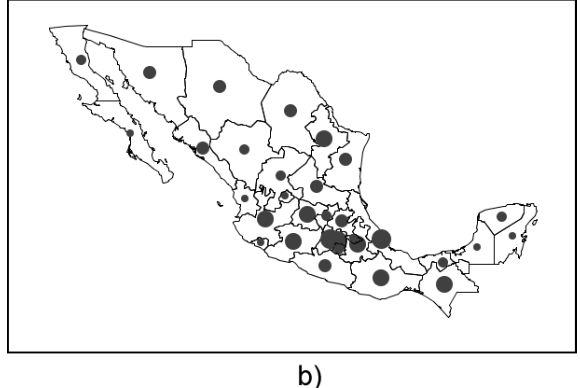


Figura 28.2: Comparación entre una representación incorrecta de la información por no emplear un esquema adecuado al tipo de esta (a) y una representación correcta utilizando un esquema coherente (b).

tes representan información cuantitativa. Esta división tiene una enorme importancia a la hora de visualizar la información temática, ya que simbolizar esta es distinto en función de sus propias características, y el uso de un esquema erróneo dará como resultado un mapa en el que no se produce una adecuada transmisión de la información. Escoger la forma adecuada de efectuar esa simbolización garantizará que los elementos visuales comunican de la mejor forma posible toda la información a la que hacen referencia. Esto puede verse claramente en el ejemplo mostrado en la figura 28.2.

Los mapas de la figura representan en ambos casos la población de los distintos estados de México, pero en cada uno de ellos se emplea una forma distinta de simbolizar los valores de población. En el primero de ellos (caso a) se ha dividido la población en cinco clases, cada una de las cuales se identifica mediante un símbolo. Los símbolos han sido escogidos de forma arbitraria, y no existe una relación entre ellos. Por su parte, el ejemplo b) también emplea símbolos y presenta igualmente cinco clases, pero en este caso tienen todos las misma forma, y lo que varía es el tamaño. Se puede establecer una relación entre los símbolos, ya que estos pueden ordenarse en función de su tamaño.

Siendo la población una variable que también puede ordenarse, el caso b) es claramente más adecuado, ya que nos proporciona la información visual de forma más rápida e inmediata. No solo responde a la pregunta *¿qué población tiene esta provincia?*, sino también a otras como *¿dónde está la provincia más poblada?* En el caso a) podemos conocer también la población de una provincia y si esta es mayor que la de otra, pero necesitamos para ello acudir a la leyenda, ya que no resulta obvio que el símbolo cuadrado indique más población que el símbolo círculo. Por su parte, el uso de un único símbolo y la variable visual tamaño es mucho más intuitivo, y nos transmite esa información sin necesidad de consultar la leyenda del mapa. Este hecho está directamente relacionado con las propiedades de las variables visuales, que ya estudiamos en el capítulo 27.

Como argumenta [328], el primer mapa es una mapa que debemos *leer*, mientras que el segundo es un mapa que podemos *ver*. Puesto que un mapa es un elemento visual, es preferible que

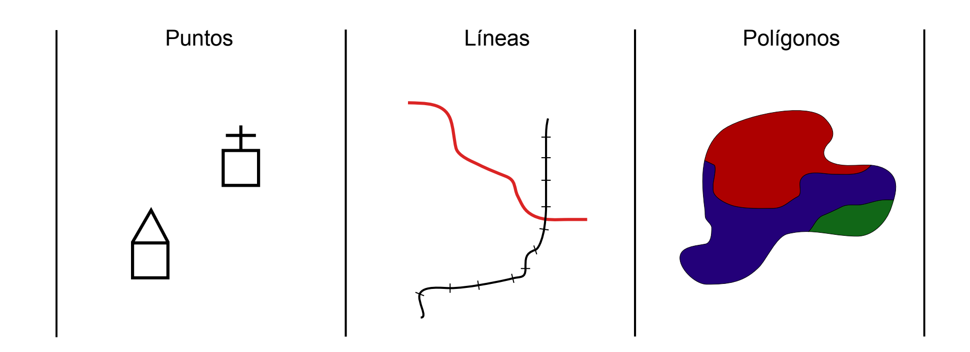


Figura 28.3: Representación de la información nominal para los distintos tipos de elementos geométricos.

transmita de forma visual su información, y un mapa a *leer* supone un desperdicio tanto de tiempo como de información misma.

Así pues, la selección de una forma de simbolización adecuada en función de la naturaleza de la información es clave para lograr un mapa efectivo. En particular, debe emplearse una variable visual que presente la propiedad (nivel de organización) adecuado. Las propiedades asociativa y selectiva solo son de interés para información cualitativa, mientras que, por ejemplo, el tamaño es la única variable visual con la propiedad cuantitativa, y por tanto la única adecuada para representar razones.

Las siguientes son algunas ideas básicas a este respecto referidas a los distintos tipos antes citados.

- **Nominal**. La información de tipo nominal se representa adecuadamente utilizando la variable visual forma. Lo que representamos responde principalmente a la pregunta *qué* en lugar de a la pregunta *cuánto*, y está más relacionado en cierto modo con la cartografía base que con la cartografía temática. El uso de símbolos, es decir, de la variable visual forma, para elementos puntuales o lineales es una solución muy eficaz y habitual en este caso. Para el caso de representar áreas puede emplearse la variable visual color y emplear distintos tonos, o bien la textura (Figura 28.3). Como dijimos en su momento, los tonos no presentan un orden (aunque citamos que pueden hacerlo si existe alguna lógica en la sucesión de estos), pero este no es necesario para este tipo de variables. La única propiedad que es de interés en este caso es la selectiva.
La información alfanumérica se trata a efectos de representación del mismo modo que la de tipo nominal.
- **Ordinal**. A diferencia de la información nominal, en la información ordinal los valores definen un orden, por lo que la propiedad ordenada es necesaria para poder aplicarla a este caso.
- **Intervalos y razones**. Tanto intervalos como razones son tipos de información con más posibilidades que las anteriores, y en las que el número de valores que encontramos a la hora de representar un fenómeno es habitualmente más elevado. Frecuentemente, estos valores son de tipo real (no enteros), por lo que es además necesario agruparlos en clases, como veremos en un próximo apartado. Como en el caso anterior, pueden emplearse todas las variables visuales que presenten la propiedad ordenada. No debe olvidarse, no obstante, que la propiedad de mostrar el orden en términos de cantidades o proporciones, que denominábamos cuantitativa, es exclusiva del tamaño, siendo este la variable visual más adecuada para representar correctamente este

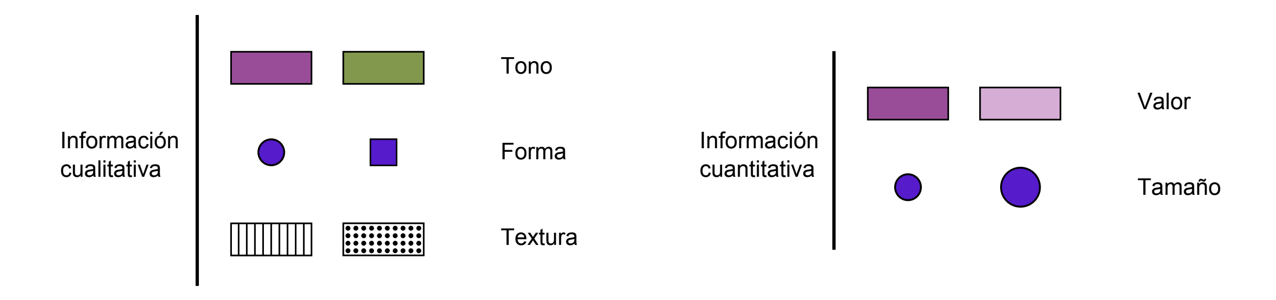


Figura 28.4: Utilización de las variables visuales según el tipo de información.

tipo de información y que al visualizar el símbolo correspondiente pueda estimarse el valor representado de forma intuitiva.

En resumen, podemos condensar este apartado con una rápida «receta» de aplicación general (aunque siempre con excepciones, ya que la representación y simbolización contiene, no olvidemos, elementos subjetivos), según los siguientes puntos:

- Para las variables cualitativas se emplean las variables visuales color, forma y textura, en la medida que sea posible según el tipo de objeto geométrico a simbolizar.
- Para las variables cuantitativas, el valor del color y el tamaño son las más adecuadas, siendo esta última la única que permite transmitir toda la información en el caso de variables de tipo razones. El tono de color puede emplearse, pero debe escogerse una gama de tonos que presente algún tipo de lógica que permita establecer un orden.

En la figura 28.4 se muestra un cuadro con estas breves ideas.

Por último, es de interés señalar que, aunque los niveles de organización de las variables visuales expresan a su vez unas posibilidades crecientes (es decir, con una variable como el valor o el tamaño podemos expresar todo lo que el tono puede transmitir, ya que están en un nivel superior), ello no implica necesariamente que el uso de una variable de un nivel superior es mejor que otra de uno inferior. Podemos ver esto claramente en la figura 28.5. En ella se ha utilizado la variable valor para representar un mapa con información cualitativa. Puesto que el valor tiene la propiedad ordenada, esto puede inducir a pensar que existe algún orden en la variable representada (tipos de suelo en este caso). Además, y debido a que el valor es disociativo, algunos elementos son más llamativos, lo que puede asociar una falsa preponderancia a la clase a la que representan.

Razonamientos similares se pueden aplicar para el caso particular de capas con variables de tipo verdadero/falso. En estas, deben emplearse colores de similares características, de forma que no exista posibilidad de interpretarlos erróneamente y asociar a alguna de las opciones la idea de ser «mejor» que la contraria. Transmitir la información no es lo único que se busca, sino también hacerlo sin que aparezcan posibles sesgos a la hora de interpretarla.

28.4.1. Creación y asignación de clases

En el caso de trabajar con información de tipo intervalos o razones, simbolizar cada uno de los valores de una forma distinta supone la

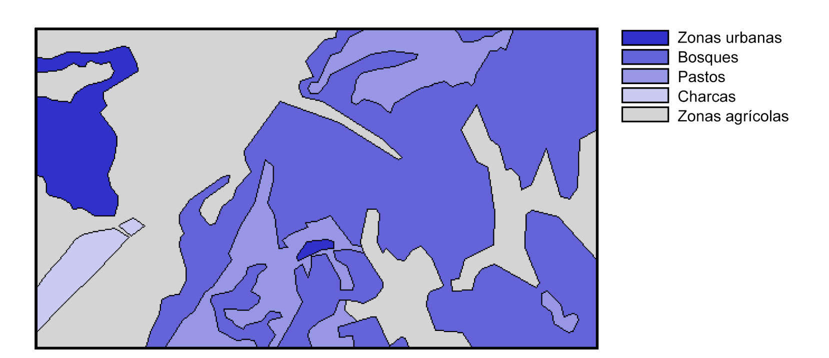


Figura 28.5: Uso incorrecto de la variable visual valor para representar información cualitativa. Puede transmitirse una falsa sensación de que existe un orden en las clases representadas.

necesidad de emplear un número muy elevado de simbologías distintas. Esto puede complicar la interpretación del mapa, especialmente si se lee este junto a su leyenda correspondiente, ya que identificar una simbología concreta en esta es complejo y resulta fácil equivocarse. Asimismo, con un número elevado de simbologías, las diferencias entre estas son pequeñas, por lo que también es complicado separar unas de otras y percibir si dos de ellas son distintas o son la misma. Por esta razón, lo habitual es agrupar todo el conjunto de valores disponibles en una serie de categorías, clasificándolos y estableciendo la simbología no en función del valor en sí, sino de la clase a la que pertenece.

La creación de clases para una serie de valores es un problema en el que han de considerarse dos parámetros principales: el número de clases a crear y el criterio a aplicar para establecer los límites de cada una.

Respecto al numero de clases, este debe ser lo suficientemente grande como para no resumir en exceso la información y poder mostrar con un cierto detalle el comportamiento de la variable, pero no demasiado alto para evitar los problemas que aparecían en el caso de no dividir los valores en clases. El número de clases es también función de la variable visual utilizada, ya que algunas resultan más fáciles de diferenciar. En general, el máximo de clases que se distinguen es del orden de 7 u 8, no siendo recomendable establecer un número mayor, con independencia de qué variable empleemos. Esto no quiere decir que deban crearse sistemáticamente 8 clases para cualquier variable y situación, ya que, en función de otros factores, puede resultar de interés elegir un número distinto. De nuevo, no debe perderse de vista la finalidad que va a tener el mapa que estamos diseñando.

Una vez que hemos decidido el número de clases, debemos definir el rango de valores que cubrirá cada una de ellas. Esto debe llevarse a cabo tratando de maximizar la información que se transmite y de aprovechar lo mejor posible la variable visual empleada. Por ejemplo, si esta variable es el valor de un color, debemos tratar que aparezca bien distribuida y que todas las clases tengan un número similar de elementos, para que todos esos valores aparezcan representados en una cantidad similar a lo largo del mapa[1].

La conveniencia de usar una u otra definición de clases está, como resulta fácil deducir, ligada

[1]Aunque en un ámbito distinto, si repasas el apartado 17.3.1 dedicado a la expansión de contraste en imágenes, encontrarás una idea similar a esta.

a la propia distribución de los valores de la variable, por lo que estudiar estos es fundamental. Un histograma es una herramienta muy útil para llevar esto a cabo.

De entre los métodos que se emplean frecuentemente para la creación de clases de forma sistemática, cabe destacar los siguientes:

- **Intervalos iguales**. Simplemente se divide el rango cubierto por los valores en n clases de la misma amplitud, siendo esta igual a $\frac{\max - \min}{n}$. Su principal inconveniente es que puede resultar en clases con muchos elementos y otras prácticamente vacías, en especial si la variable tiene una distribución normal o aparecen elementos con valores atípicos (*outliers*), que desvirtúan el significado del máximo y el mínimo a la hora de calcular la amplitud de cada clase.
- **Intervalos naturales**. Basados en la propuesta de *saltos naturales* de Jenks [331], trata de establecer clases lo más homogéneas posibles, disminuyendo la varianza de cada clase. De este modo, se obtienen clases que presentan la máxima variabilidad entre ellas, constituyendo categorías bien diferenciadas unas de otras.
- **Intervalos normales**. De especial interés para el caso en que la variable presenta una distribución normal. Se toma la media de los valores y se crean los límites de cada clase sumando o restando a esta la desviación típica o un múltiplo de esta.
- **Intervalos por percentiles**. Utilizando percentiles pueden crearse clases de tal modo que todas ellas contengan el mismo número de elementos. Por ejemplo, los *cuartiles* dividirán el rango de valores en cuatro clases, cada una de ella con igual numero de elementos. En este caso, los límites de separación de clases se encontraran en los percentiles del 25, 50 y 75 por cien, respectivamente.

  Pueden aplicarse también los percentiles no sobre la variable que se representa, sino sobre la superficie que ocupan sus distintos valores. Se tiene de este modo los *percentiles de superficie*, que crean n clases, todas ellas representadas en el mapa por una misma superficie.
- **Intervalos en progresión**. Pueden emplearse progresiones como la aritmética o la geométrica para crear las clases, en caso de que los valores de la variable a representar muestren un comportamiento según alguna de estas progresiones.

Una comparación visual del resultado de aplicar algunos de los métodos anteriores se muestra en la figura 28.6

Junto a lo anterior, pueden utilizarse transformaciones de los valores previas a su asignación a una clase, para después clasificar el valor transformado. Una transformación logarítmica es habitual para el caso de valores distribuidos irregularmente, con muchos de ellos en un rango dado y unos pocos en un rango alejado de este. Aplicando un logaritmo (generalmente de base 10), los valores transformados pueden mostrar, por ejemplo, una distribución normal, siendo entonces posible aplicarles una simbolización mediante intervalos normales. Vimos un ejemplo de esto en la figura 15.4.

Aunque resulta práctico definir las clases utilizando alguna de las metodologías anteriores, pueden igualmente establecerse límites de clase arbitrariamente según se considere oportuno en función de la distribución de los valores. Por ejemplo, si existen saltos importantes en esta y quiere reseñarse este hecho, pueden incluirse explícitamente como límites de los intervalos. Asimismo, pueden incorporarse valores particulares que sean de importancia para la variable representada. Esto puede verse claramente en el ejemplo de la figura 28.7

Para el caso mostrado, en el cual se representa la elevación, es interesante diferenciar los valores positivos (sobre el nivel del mar) de los

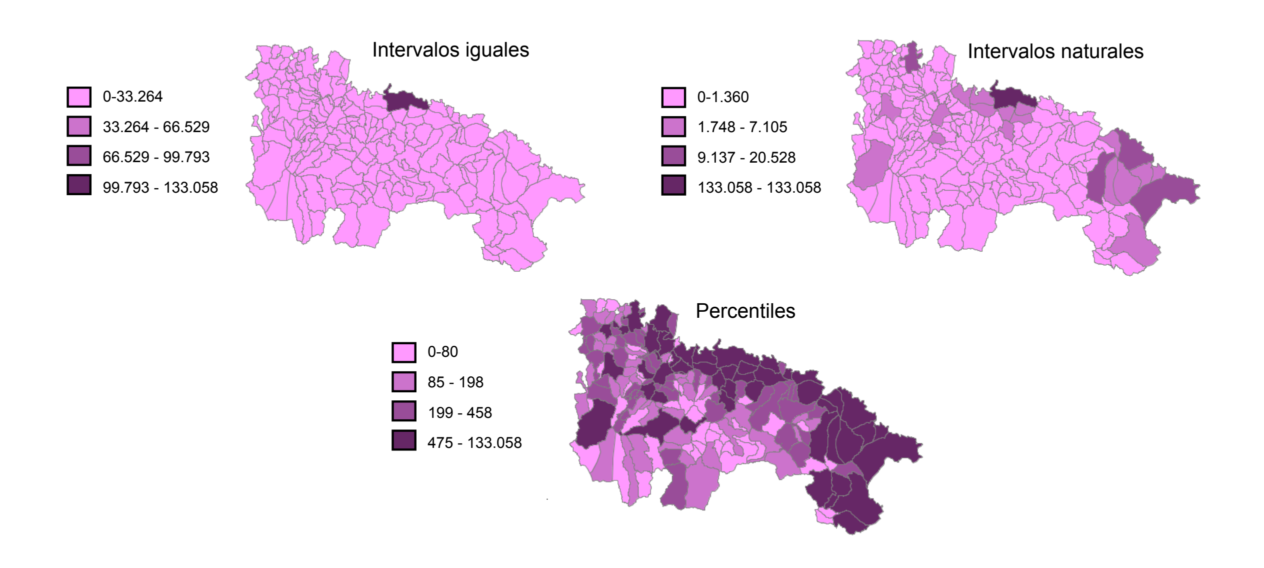


Figura 28.6: Comparación entre distintos esquemas para la creación de intervalos de clase.

negativos (zonas por debajo del nivel del mar y, especialmente, batimetría del fondo marino). El cero es un valor que puede o no aparecer de modo natural como límite de clase al analizar los datos de elevación, pero que se incorpora por su importancia.

El mapa de la figura presenta además un caso particular por otras razones, ya que utiliza el color como variable ordenada, pese a que dijimos que normalmente no posee tal propiedad. No obstante, este es uno de esos casos en que sí existe un orden fácil de percibir, ya que los colores escogidos están pensados para ser identificados con distintas zonas altitudinales. Las zonas de batimetría se representan en tonos de azul, por lo que en ese tramo se está empleando realmente la componente del color que denominábamos valor. Para las restantes, se comienza en el verde (zonas bajas donde crece vegetación que es de ese color), seguido del marrón (zonas altas sin vegetación) y después el blanco (zonas elevadas que se pueden asociar a nieve). La división en esos tramos se hace empleando el valor igualmente. Esta asociación de conceptos tan básica (y no necesariamente muy real, pero sí conocida y compartida por todo el mundo) permite crear un orden y capacitar a la variable visual color para emplearse a la hora de representar una variable de tipo intervalo como es la elevación.

La presencia del valor cero como punto que define dos mitades (elevaciones sobre el nivel del mar o por debajo de este) hace que los datos de elevación tengan, en lo que a su simbolización respecta, un esquema de tipo *divergente*. Este tipo de esquemas aparecen cuando la variable presenta algún valor crítico con un significado particular, dividiendo el conjunto de valores en grupos que pueden considerarse independientes. Es habitual emplear un color de valor bajo (esto es, un color claro) en las cercanías del punto crítico, y aumentar el valor a medida que nos acercamos a los extremos tanto por encima como por debajo de este punto. Cada mitad, a su vez, suele representarse con colores que presentan un fuer-

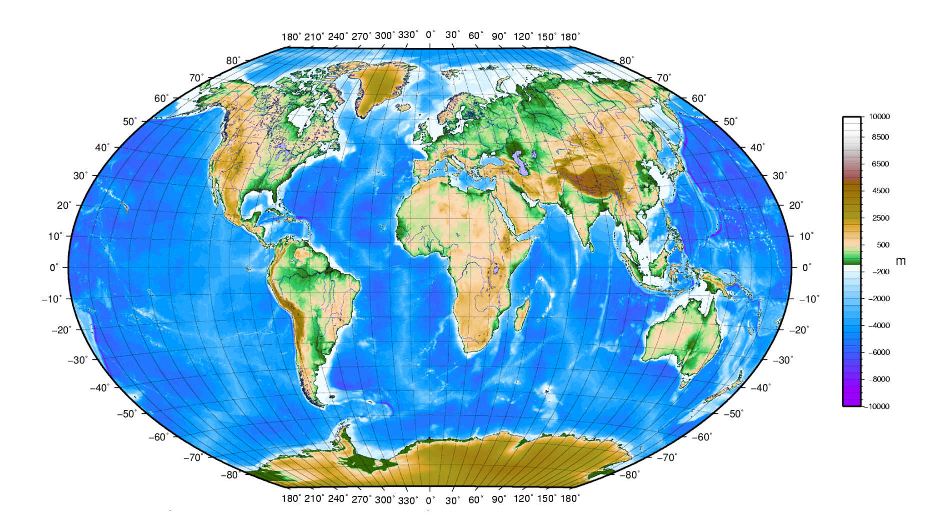


Figura 28.7: Los intervalos pueden incorporar valores de importancia para una determinada variable. En este caso, para la variable elevación resulta particularmente relevante el valor cero, que delimita el comienzo de las clases representadas en azul.

te contraste entre sí, para de este modo indicar que cada una de ellas representa una realidad bien distinta de la otra.

Los esquemas no divergentes para variables cualitativas se dice que son de tipo *secuencial*.

Debe reseñarse que, en el caso de establecer las clases en función de los datos, tal y como sucede al aplicar los métodos que hemos descrito, la simbolización no será adecuada para realizar comparaciones con otros mapas. Un mismo valor puede simbolizarse con colores distintos en dos mapas dados, ya que la clase a la que pertenece depende del resto de valores en su conjunto, por lo que no tiene sentido una comparación visual. Por el contrario, si el intervalo se define sin considerar los valores particulares del conjunto representado (como en el mapa de elevaciones anterior), el mismo color en dos mapas sí que implica un mismo rango de valores, con lo que pueden efectuarse comparaciones.

Si quieres experimentar con la definición de clases y la asignación de colores a estas, una herramienta de enorme valor es la que encontrarás en la pagina Web `http://www.colorbrewer.org`. Úsala no solo para probar ahora todo lo explicado en este capítulo, sino también cuando tengas que crear tus propios mapas. Elegir un conjunto adecuado de colores y clases no es una tarea sencilla, y una herramienta así puede aportar mucho valor a tus mapas si la empleas correctamente junto a las propias funcionalidades del SIG que estés utilizando.

28.5. Elementos del mapa. Composición

Un mapa no es solo una colección de gráficos que representan objetos o valores del mundo real a una escala dada, sino que para ser verdaderamente completo requiere completarse con otra serie de elementos adicionales. Es decir, el mapa en sí no es solo lo que se deriva de la representación de la información geográfica y su simbolización, sino un conjunto de elementos dis-

puestos de forma óptima, entre los cuales, eso sí, resulta de particular relevancia aquel que contiene la información geográfica como tal.

Igual de importante que simbolizar de manera correcta la información geográfica es situar adecuadamente los distintos elementos del mapa, ya que estos están pensados también, al igual que la propia simbología, para facilitar la interpretación de la información y hacer esta más comprensible.

Los siguientes son los elementos fundamentales que podemos emplear para componer un mapa (Figura 28.8):

- **Nombre o título**. Imprescindible para conocer qué información muestra el mapa.
- **Autor**. La persona u organismo que ha creado el mapa debe aparecer indicada en algún punto de este.
- **Otra información sobre el mapa**. Por ejemplo, la relativa al sistema de referencia empleado o la fecha de su creación, entre otras.
- **Canevás**. El canevás nos indica dónde dentro de la superficie terrestre se encuentra aquello que el mapa representa, y provee la referencia geográfica para sus elementos. Asimismo, complementa a la escala para la estimación visual de distancias y medidas. Es más necesario en caso de escalas bajas, aunque se añade con independencia de la escala.
- **Leyenda**. Aunque se ha de tratar de utilizar una simbología lo más expresiva posible, no toda la información puede incorporarse en el mapa, y es necesario acompañarlo de una leyenda. Esta ha de ser también fácil de interpretar y lo más clara posible. Una leyenda demasiado extensa o de difícil comprensión probablemente nos indica que la simbología escogida es mejorable.
  La leyenda y el mapa en sí forman un todo, por lo que no deben separarse mediante un cuadro, salvo en el caso en que el mapa cubra todo el área del lienzo y no sea fácil separar visualmente de forma clara ambos elementos.
- **Norte**. Aunque habitualmente se presupone la orientación Norte-Sur, no siempre ha de ocurrir así, y una aguja apuntando al norte o una rosa de los vientos sirve para aclarar la orientación del mapa. Es de reseñar que la orientación no ha de ser constante para todos los puntos de un mapa, estando esto en relación con el tipo de sistema de coordenadas y la proyección empleada. Por ejemplo, en el mapa mundial de la figura 28.7, el Norte se sitúa hacia arriba de la hoja solo en el centro. Si nos encontramos en la parte izquierda del mapa la dirección del Norte no es la misma. El canevás, que contiene los paralelos y meridianos, será en este caso la referencia fiable en lo que a orientación respecta.
- **Escala**. La escala debe indicarse tanto de forma numérica como gráfica, de modo que puedan realizarse cálculos y estimar visualmente distancias entre puntos dados del mapa.
- **Localizador**. Un localizador provee un elemento visual para situar el mapa en un contexto geográfico más amplio, de modo similar al canevás. Es de especial interés en el caso de series de mapas, para establecer la relación entre el presente y los restantes dentro de la misma serie. En este caso, el localizador sirve como mapa índice.
- **Mapas de detalle**. Cuando resulta necesario mostrar una cierta zona del mapa con mayor detalle y a una escala mayor, se puede incluir un mapa correspondiente a esa zona como un enclavado dentro del mapa principal. Se debe señalar asimismo sobre este último la zona a la que corresponde el mapa de detalle.

Aunque en un mapa en sentido clásico deben incorporarse todos o la gran mayoría de los anteriores elementos, cuando trabajamos con representaciones dentro de un SIG la situación es distinta y se puede prescindir de una buena parte de ellos. Por ejemplo, y dado el carácter menos

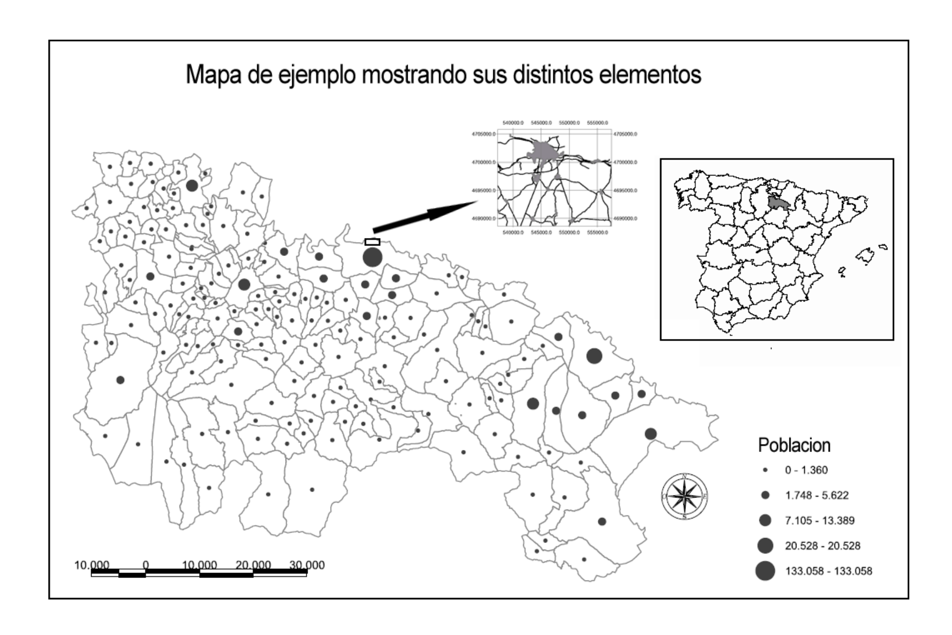


Figura 28.8: Ejemplo de mapa mostrando sus elementos más habituales.

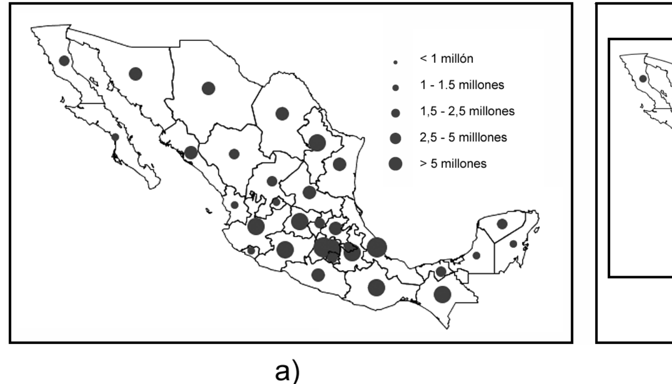
persistente de la representación en pantalla, añadir el nombre del autor carece la mayoría de las veces de sentido. Información tal como la procedencia de los datos que estamos visualizando resulta de más interés que el autor del mapa, pero lo correcto es consultar esta en los propios datos, que deberían contenerla de algún modo (vimos esto en el capítulo 9).

La escala es adecuado mostrarla de forma numérica, pero no tanto en su versión gráfica, ya que dentro de un SIG encontramos herramientas que nos permiten medir con total precisión distancias y áreas, y una escala gráfica carece de utilidad en este contexto. Por su parte, el localizador es mejor que el canevás para definir el contexto, ya que muchas aplicaciones SIG incorporan incluso un localizador interactivo sobre el que puede operarse para cambiar el encuadre del mapa.

En lo que respecta a la forma de disponer los elementos sobre el lienzo que un mapa conforma, la premisa fundamental es maximizar la claridad y aprovechar de la mejor forma posible el espacio disponible. La figura 28.9 muestra un claro ejemplo de cómo un adecuado uso del espacio en el mapa, evitando que existan zonas en blanco que no comunican ninguna información, mejora notablemente la calidad del mapa.

Asimismo, es importante que el diseño del mapa recalque su propósito, haciendo énfasis en los aspectos más relevantes para cumplir este.

Aunque el objetivo principal del diseño cartográfico es crear un mapa útil y no un mapa bonito, no cabe duda que una cierta preocupación por el aspecto estético es recomendable, ya que también contribuirá a una mejor interpretación de la información del mapa. Este es un aspecto subjetivo y con una componente principalmente artística, aunque también pueden aportarse algunos elementos metodológicos de carácter más sistemático. Uno de ellos utilizado frecuentemente es el empleo de la proporción áurea para dimensionar los elementos del mapa. Comenzando por las dimensiones del propio lienzo, puede aplicarse



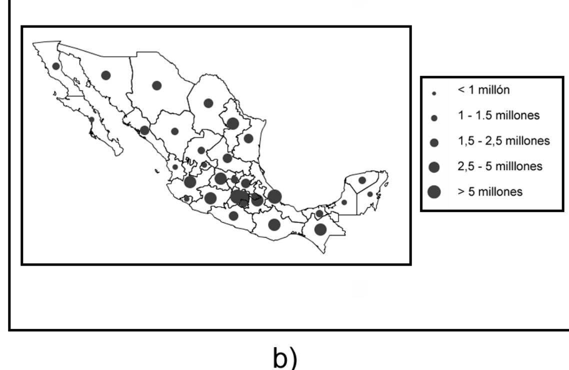


Figura 28.9: Ejemplo de un aprovechamiento óptimo del espacio de un mapa (a) y un aprovechamiento incorrecto de este (b).

a las de los restantes componentes, tales como la leyenda en caso de estar situada en un cuadro aparte, o el cuadro que contiene el nombre del mapa y otra información adicional.

Los conceptos que deben manejarse a la hora de elegir las características de los elementos del mapa y su emplazamiento derivan de la percepción visual, disciplina que ya vimos en el capítulo anterior. A continuación tienes algunas ideas adicionales sobre percepción visual que deben aplicarse a la composición de mapas. Si deseas ampliar estos conceptos, una buena referencia sobre percepción visual desde el punto de vista del arte es [332].

- El documento cartográfico tiene dos centros. Un centro geométrico y uno óptico. Este último se sitúa por encima del geométrico, aproximadamente a un 5 % de la altura total del documento. Los elementos del mapa se deben disponer alrededor del centro óptico.
- Los elementos en la parte superior del mapa tienen una mayor importancia, así como los situados en la parte izquierda. Es en estas zonas donde deben situarse los elementos más importantes sobre los que se quiera centrar la atención.
- La atención del lector del mapa va desde la esquina superior izquierda hasta la inferior derecha, pasando por el centro óptico. Los elementos importantes deben situarse en esta línea, para que su posición se corresponda con los movimientos naturales de la vista.
- Debe tratarse de crear un mapa que sea visualmente equilibrado. El equilibrio visual es el resultado del peso que cada elemento tiene y su posición, así como su orientación. Estos pesos deben repartirse adecuadamente por todo el lienzo del mapa. El peso de un elemento depende de múltiples factores, entre ellos los siguientes:
    - Posición. Los elementos tiene más peso en la derecha que en la izquierda, y más en la parte superior que en la inferior. El peso aumenta al aumentar la distancia al centro del documento.
    - Tamaño. Mayor tamaño implica más peso.
    - Color. Los colores brillantes tienen más peso que los oscuros. El tono rojo tiene más peso que el azul.
    - Aislamiento. Los elementos aislados tienen más peso que aquellos rodeados por otros.

- Forma. Las formas regulares tienen más peso que las irregulares. Cuanto más compacta sea la forma, también tendrá más peso.
- Dirección. Algunos elementos pueden tener una dirección que «dirija» la atención hacia otros, concediéndoles peso (por ejemplo, una flecha que señale a un elemento, haciendo que llame más la atención),

Las ideas acerca de la composición y el equilibrio del mapa se han de aplicar a todo el documento cartográfico (es decir, al que contiene todos los elementos citados anteriormente), así como a la parte de este que representa la información geográfica. Es importante seleccionar adecuadamente el área geográfica cubierta para que la información relevante que se muestra acerca de esta conforme un conjunto equilibrado y siga a su vez las indicaciones mencionadas.

Recordar, por último, que la composición del mapa implica una organización horizontal (plana) de sus elementos, pero existe asimismo una organización vertical. Esta viene definida por la jerarquía existente, sobre la cual ya se comentaron algunas ideas en el apartado 27.5.2. Estas ideas deben aplicarse igualmente en la composición del mapa, para conjuntamente lograr un documento equilibrado en el que quede claro qué elementos son los de mayor importancia, y en el que pueda accederse con facilidad a la información que contienen.

28.6. Tipos de mapas temáticos

Los mapas temáticos representan la mayor parte de los creados en un SIG, por lo que resulta necesario ver en detalle las formas en las que pueden presentarse. Existen diversas alternativas en función del tipo de elemento que se pretenda simbolizar o las características de la variable tratada, y la elección de una u otra supondrá una diferencia importante en el mapa obtenido y en su uso posterior. En un mismo mapa pueden combinarse varias de estas formas, especialmente si se pretende representar más de una variable, en cuyo caso la combinación debe buscar la máxima claridad en la representación de todas ellas.

En este apartado detallaremos los siguientes tipos de mapas temáticos: mapas de símbolos proporcionales, mapas de densidad de puntos, mapas de coropletas, y mapas de isolíneas. Todos ellos se utilizan para la representación de variables cuantitativas.

28.6.1. Mapas de símbolos proporcionales

Un mapa de símbolos proporcionales representa variables cuantitativas a través de símbolos cuyo tamaño esta en relación con el valor a representar de dicha variable. Es decir, emplea la variable visual tamaño, que como ya hemos visto es la única que presenta la propiedad cuantitativa. La forma de los distintos símbolos es siempre la misma, y por simplicidad lo más frecuente es utilizar como símbolo base el círculo, aunque puede utilizarse cualquier otro, e incluso símbolos de tipo lineal (barras).

Puesto que el tamaño es el elemento que diferencia a los distintos símbolos y el que transmite la información cuantitativa, su elección es crucial para la creación de un buen mapa de este tipo. La elección de un tamaño implica elegir uno mínimo y uno máximo, correspondientes a los valores mínimo y máximo de la variable en el mapa. Entre estos se situarán los distintos tamaños correspondientes al resto de posible valores que toma la variable.

Existe, claramente, una relación entre el tamaño máximo y el mínimo, ya que se define una relación de escalado de los distintos valores. Este escalado es distinto para símbolos lineales que para símbolos de área, ya que la percepción de la

relación entre ellos es distinto según el tipo de símbolo empleado. En ambos casos, el escalado debe ser coherente con el valor que se representa, de tal modo que si el usuario del mapa percibe que el tamaño de un símbolo es el doble que el de otro, los valores de ambos símbolos estén igualmente en esa proporción.

Para conseguir esto se ha de seleccionar el tamaño asociado al valor de uno de los extremos. Esto se hará con un criterio puramente gráfico, de tal modo que, si por ejemplo establecemos el tamaño máximo, este no sea excesivo y a la hora de representar el símbolo correspondiente en el mapa ocupe demasiado espacio y existan solapes. Debe evitarse asimismo que el tamaño mínimo sea demasiado pequeño y no se aprecie el símbolo con claridad. Una vez hecho esto, se establece una relación lineal, de tal forma que podemos calcular el tamaño correspondiente a todo valor. Si un valor de 100 se corresponde con una barra de una altura de 10 mm, entonces un valor de 200 se representara mediante una barra de 20 mm, y así sucesivamente.

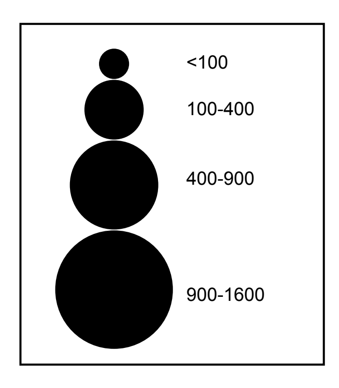
Para el caso de símbolos superficiales, no obstante, el escalado no debe hacerse en función de un parámetro lineal (por ejemplo, el radio en el caso de emplear círculos), sino respecto a la propia superficie. Es decir, si un valor de 100 se representa con un circulo de radio r, el valor 200 no se representará mediante un círculo de radio $r' = 2r$, sino con una de tal radio que la superficie sea el doble del primero. En este caso, el radio buscado sería $r' = \sqrt{2}r$.

El escalado de símbolos se puede dar de forma continua, de tal modo que cada valor se representa con un símbolo de un tamaño calculado según la idea anterior, empleando el valor exacto para el escalado. No obstante, la capacidad de diferenciar visualmente tamaños distintos e interpretar la relación entre ellos es limitada, por lo que suele resultar más conveniente efectuar un escalado discreto. Es decir, crear clases y asignar a un valor no un símbolo del tamaño exacto que le correspondería, sino el asignado al valor que define a la clase, habitualmente el centro de esta.

Tanto las barras como los círculos pueden sectorizarse, mostrando una división en subclases del valor total que representan. Para el caso de la población, podrían mostrarse las proporciones que corresponden a hombres y mujeres. Este tipo de representaciones, no obstante, son a veces difíciles de interpretar en su conjunto, por lo que resulta más adecuado crear varios mapas que muestren esa misma información por separado, en lugar de conjuntamente en uno único.

Aunque la variable visual tamaño presenta la propiedad cuantitativa, la percepción de la relación de tamaño no es perfecta y existe una cierta imprecisión. Esta se debe a muchos factores, como por ejemplo el hecho de que los símbolos situados alrededor de uno dado pueden afectar a la percepción de su tamaño. Por esta razón, es importante para facilitar la correcta interpretación de un mapa de símbolos graduados el mostrar en la leyenda la relación entre los distintos tamaños de los símbolos y sus valores. Para el caso habitual de emplear círculos, esto puede llevarse a cabo mediante elementos gráficos como los mostrados en la figura 28.10

El uso de un escalado lineal en el que se conserve la propiedad cuantitativa resulta en ocasiones inapropiado debido a la distribución de los valores. Por ejemplo, para representar el mapa de la figura 28.6, este esquema no es adecuado, ya que una de las zonas presenta un valor de la variable muy superior a la del resto (puede verse esto claramente en la representación por intervalos iguales), lo cual requeriría el uso de un símbolo desproporcionadamente grande. Si se usan clases iguales, la mayoría de los valores entrarían en una de ellas, por lo que no se transmitiría bien la distribución de estos. En este caso, se debe em-



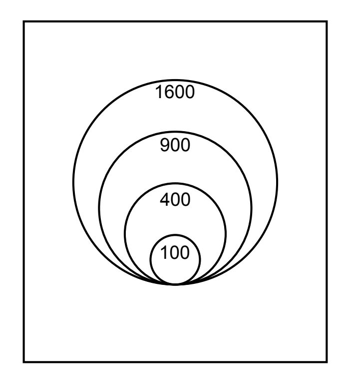


Figura 28.10: Dos ejemplos de leyendas para un mapa de símbolos proporcionales.

plear un esquema de clases distinto, aunque así la proporción de tamaños no permita visualmente estimar las cantidades. Es decir, los tamaños de los símbolos nos indican que hay más cantidad en una zona que en otra, pero no podemos solo con ellos saber *cuánto* más hay. Los mapas elaborados de esta forma se conocen como mapa de *símbolos graduados*. En estos mapas, la importancia de la leyenda es aún mayor si cabe, ya que es la encargada de explicar el significado de cada tamaño, y sin ella la información de la que disponemos es mucho menor.

El mapa de la figura 28.8, que mostramos al presentar los distintos elementos del mapa, es un ejemplo mapa de símbolos graduados.

28.6.2. Mapas de puntos

Los mapas de puntos se emplean especialmente para la representación de variables que representen algún tipo de cantidad, tales como la población, el gasto medio por persona o la producción de un determinado cultivo. Estas cantidades se representan mediante la repetición de puntos, en numero proporcional a su magnitud. Cada uno de esos puntos representa un valor unitario, y el conjunto de ellos sobre la zona en cuestión suma la cantidad total a representar. Los puntos tienen todos la misma forma y tamaño, a diferencia de lo que vimos en el caso de los símbolos proporcionales.

Los mapas de puntos transmiten de forma muy eficaz los valores que representan, obteniéndose estos por el mero recuento, aunque visualmente permiten una estimación inmediata y pueden compararse entre las distintas zonas del mapa. Por esta razón, son especialmente adecuados para variables discretas más que para continuas, aunque también pueden emplearse para estas últimas.

Aunque podrían crearse con cualquier otro símbolo, ya que es la repetición de este la que transmite la información, lo más habitual es el empleo de puntos, de ahí el nombre genérico que se les da.

Tres son los aspectos que deben tenerse en cuenta a la hora de elaborar un mapa de puntos: el valor de cada punto (es decir, cuántas unidades de la variable representa cada punto), su tamaño y su posición.

Si los valores de la variable que se manejan son bajos, se puede establecer como valor del punto la unidad. Es decir, un punto representa sobre el mapa un habitante en el caso de un mapa de población. No obstante, con valores altos (como en el caso de la población) esto da lugar a un número demasiado elevado de puntos que

saturan el espacio del mapa y no transmiten adecuadamente la información. Por ello, cada punto debe representar un número mayor de elementos de la variable representada, de tal modo que no aparezcan en demasía en el mapa, solapándose unos con otros.

Si el valor escogido es demasiado alto, aparecerán pocos puntos en el mapa, y este puede quedar poco expresivo y no transmitir la distribución de la variable. Debe, por tanto, escogerse un valor adecuado que equilibre la presentación de los puntos sobre el mapa. Este valor se representará en la leyenda para su interpretación, habitualmente en forma de texto, escribiendo por ejemplo, que «un punto equivale a 1000 habitantes».

La elección del tamaño del punto debe garantizar la buena visibilidad de este, al tiempo que no debe ser excesivamente grande para que no ocupe demasiado espacio y dificulte la visión de otros. Obviamente, el tamaño óptimo está en relación con el valor unitario escogido, y ambos parámetros deben establecerse conjuntamente para lograr la combinación más adecuada.

Por último, la posición del punto es de gran importancia para transmitir la información correcta y no dar lugar ambigüedades o incorporar errores conceptuales. Si no disponemos de información adicional y solo tenemos el valor correspondiente a una zona dada, los puntos se han de disponer de forma regular ocupando toda la superficie de la zona. Si, por el contrario, sabemos algo más acerca de la distribución de la variable, debemos emplear esa información para emplazarlos de forma más realista. Si, por ejemplo, la zona corresponde a una provincia y sabemos la localización de la principal ciudad dentro de ella, es más lógico situar más puntos cerca del emplazamiento de esa ciudad que en otras partes de la provincia, ya que una mayor parte de la población estará allí.

Otro aspecto a considerar es el significado de la variable que se representa y la posibilidad o no de que aparezca en las distintas localizaciones de los puntos. Si la variable es, por ejemplo, el numero de ejemplares avistados de un determinado ave acuática, situar los puntos sobre zonas urbanas o de bosque no tiene sentido, ya que dan a entender que ahí hay presencia de esa especie (tantos ejemplares como los puntos en cuestión indiquen), algo que es falso.

En los dos casos anteriores va a resultar necesario «mover» los puntos a su localización más correcta, algo que, habitualmente, no resulta posible con los mecanismos automatizados de que dispone un SIG. El chequeo del mapa creado resulta, por tanto, imprescindible para comprobar que existen puntos en posiciones erróneas. El uso de herramientas externas tales como programas de diseño gráfico, según vimos en el capítulo 26, es una solución para retocar los mapas creados y obtener una distribución de los puntos más correcta.

La imagen 28.11 muestra un ejemplo de un mapa de puntos.

28.6.3. Mapas de isolíneas

Los mapas de isolíneas están entre lo más usados para la representación de información cuantitativa, en particular cuando se trata de variables continuas. Se utilizan habitualmente para representar campos escalares, y constituyen una forma muy efectiva de incorporar esta información en un mapa, ya que pueden combinarse con otros tipos de mapas y de información, debido a que, al representarse únicamente mediante líneas, permite la presencia de otros elementos dentro del mapa sin resultar estas obstrusivas.

Un mapa de isolíneas está formado por un conjunto de líneas, cada una de las cuales une puntos que presentan el mismo valor de la va-



Figura 28.11: Mapa de puntos.

riable. Estas líneas no pueden cruzarse, ya que ello significaría que en un punto se presentan dos valores. El caso más típico de mapa de isolíneas son las curvas de nivel que aparecen el un mapa topográfico, indicando la elevación del terreno. Otras variables que habitualmente se representan mediante curvas de nivel son la temperatura (en cuyo caso, las líneas se denominan *isotermas*), la presión (*isobaras*) o el tiempo (*isocronas*). En el caso de las curvas de elevación, estas se conocen como *isohipsas*, aunque resulta mucho más habitual denominarlas simplemente curvas de nivel, nombre que se emplea también por extensión como sinónimo general de isolíneas.

Para una variable continua, los valores que esta puede tomar son infinitos, por lo que el número de isolíneas que pueden trazarse también lo es. Por ello, es necesario seleccionar qué isolíneas se desea representar, estableciendo clases y representando tan solo los límites de estas. A pesar de esta división, no resulta habitual un análisis complejo a la hora de establecer la distintas clases, tal y como se detalló en el apartado 28.4.1. En su lugar, se emplean en la gran mayoría de casos intervalos iguales, siendo el tamaño de cada clase (el rango de valores que cubre) el único parámetro a definir. Este parámetro es lo que se conoce como *equidistancia* en un mapa de curvas de nivel.

La construcción de un mapa de curvas de nivel implica estimar el trazado de estas a partir de valores puntuales, lo cual coincide con lo que vimos en el capítulo 14 acerca de los distintos métodos de interpolación. Dentro de un SIG, el procedimiento a seguir será calcular una capa ráster a partir de valores puntuales, y después crear las isolíneas a partir de esta capa según lo visto en el apartado 18.3, para una equidistancia dada.

A la hora de simbolizar las isolíneas, y con independencia de su tipo, la variable visual tamaño es la única que suele emplearse, en particular para señalar aquellas líneas que representan un valor múltiplo de una determinada cantidad y hacer así más fácil la lectura del mapa. Estas líneas son lo que se conoce como *curvas directrices*. Por ejemplo, en un mapa topográfico con curvas de nivel con una equidistancia de 100 metros, es habitual establecer curvas directrices cada 500 metros. Todas aquellas curvas cuyo valor asociado sea múltiplo de 500 se representan con un

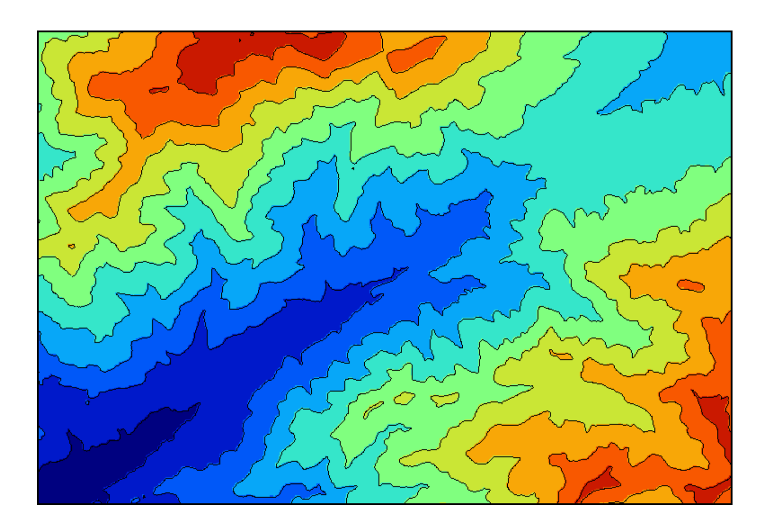


Figura 28.12: Mapa de isocoropletas.

trazo más grueso para que puedan localizarse rápidamente.

El uso del color o la textura en las líneas no es habitual como simbología, ya que simbolizar los valores de cada una a través de las variables visuales resulta en este caso menos práctico. Lo normal es etiquetar cada una de ellas con el valor concreto (con texto sobre la línea), y aprovechar el hecho de que dos líneas consecutivas están separadas siempre una magnitud igual al tamaño de la clase (la equidistancia), lo cual aporta un importante contexto en lo que a los valores se refiere.

Una forma particular de representar las isolíneas mediante color es hacerlo no sobre las líneas, sino sobre las zonas que median entre ellas. Es decir, representar la clase en lugar del límite de clase. Este tipo de mapas se asemeja al mapa de coropletas (que veremos seguidamente), tratándose más de un mapa de áreas que de líneas, por lo que se conoce como de *isocoropletas*. Ambos tipos de representación, mediante áreas y mediante líneas, pueden combinarse en un único mapa.

En la figura 28.12 puede verse un ejemplo de mapa de isocoropletas.

28.6.4. Mapas de coropletas

Los mapas de coropletas se utilizan de manera muy habitual en un SIG, y hemos visto ejemplos de ellos en otros puntos de este y otros capítulos. Por ejemplo, los mapas de la figura 28.6 son todos ellos mapas de coropletas.

En un mapa de coropletas se tiene una serie de áreas definidas, cada una de las cuales posee un valor de una variable. Este valor de la variable afecta a todo el área y es el que se representa por medio de alguna variable visual, normalmente el color a través de su componente valor. Las zonas definidas por cada área tienen un significado arbitrario, no relacionado con la variable asociada. Muy frecuentemente, se utilizan limites administrativos o de gestión para este propósito. Cada área conforma una unidad espacial, y el valor asociado a ella resume la variable dentro de dicho área.

Precisamente por esta generalización que se da al representar mediante un único valor la va-

riable dentro de cada unidad, los mapas de coropletas adolecen de ciertos inconvenientes, siendo los dos siguientes los principales:

- **Sensación de cambio brusco en los límites entre áreas**. Al existir una transición abrupta entre unidades, un mapa de coropletas puede transmitir la idea de que en esa frontera los valores de la variable cambian bruscamente, ocultando la continuidad de la variable en caso de existir esta.
- **Homogeneidad dentro de cada área**. La variación dentro de cada área no se recoge, con lo que se pierde una parte de la información. El uso de unidades menores soluciona en parte este problema, aunque puede hacer el mapa más complejo de interpretar y puede desvirtuar la información (recordemos aquí todo lo que vimos en el capítulo 11 y los conceptos tales como el Problema de la Unidad de Área Modificable). Al mismo tiempo, las unidades pueden tener su significado particular, como por ejemplo tratarse de divisiones administrativas, con lo que el uso de otras distintas altera la información que se pretende transmitir.

Igualmente, debe considerarse que, en el caso de valores no normalizados, las coropletas pueden transmitir una información equivocada. Por ejemplo, si una variable representa un conteo, tal y como la población de un conjunto de estados, el uso de coropletas no tiene en cuenta la superficie de cada una de las áreas representadas. Un mismo valor en dos unidades, una de ellas con una superficie mucho mayor a la otra, puede dar la sensación de que poblacionalmente ambas zonas son similares, mientras que puede ser que una tenga una gran densidad de población y la otra esté prácticamente despoblada. El valor que simbolizamos sí está relacionado con el área (a mayor área, encontraremos más habitantes), y sería más adecuado representar esa densidad de población, ya que resulta menos proclive a inducir una interpretación errónea. En general, el uso de coropletas es correcto cuando la variable ha sido normalizada, por ejemplo dividiendo el valor numérico de cada unidad entre la superficie de esta.

En los mapas de coropletas cobra especial importancia la correcta división de clases según hemos detallado dentro de este mismo capítulo. De entre las variables visuales, el color es la usada en la gran mayoría de casos, en particular utilizando su componente valor, y las propias características de las coropletas, en particular las desventajas que ya hemos mencionado, han de considerarse a la hora de establecer cómo hacemos uso de esta variable visual para la simbolización de cada unidad.

Así, debemos tener en cuenta que a la hora de distinguir dos colores con el mismo tono y distinto valor, si estos son muy semejantes solo resulta posible diferenciarlos cuando se sitúan el uno junto al otro, pero no cuando están separados y median entre ellos otros colores distintos. Aunque la variable con la que trabajemos sea continua, el mapa de coropletas no ha de exhibir dicha continuidad, por lo que no podemos contar con ella para elaborar la rampa de valores correspondiente. En un mapa de coropletas, una unidad puede tener a su lado otra con un valor muy distinto sin que entre ellas exista una de valor intermedio, pudiendo producirse un salto de varias clases. Esto tiene como consecuencia que el número de clases que podemos emplear es más reducida, ya que esta separación espacial que puede aparecer en las distintas clases va a dificultar su diferenciación.

De igual modo, el uso del tono queda más restringido, al poder dar lugar a situaciones ambiguas. Por ejemplo, si miramos la leyenda del mapa de la figura 28.7 veremos que hay dos clases con un tono blanco. Por una parte, los valores situados cerca del cero (al nivel del mar).

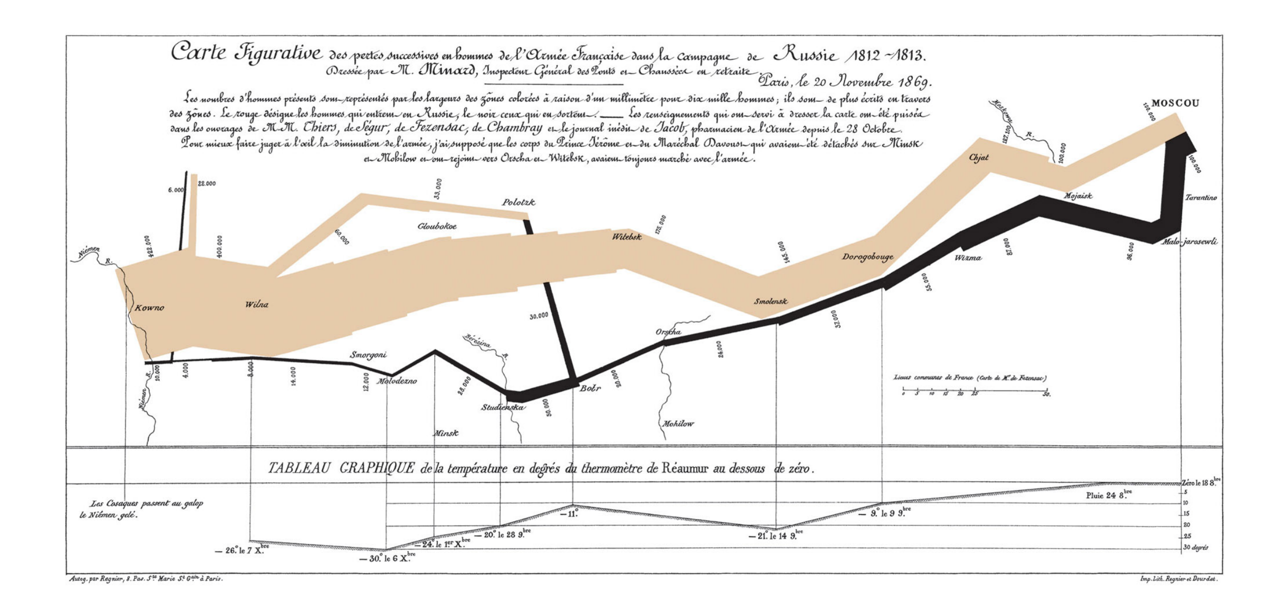


Figura 28.13: Mapa de flujo de Charles Joseph Minard sobre la campaña de Napoleón en Rusia.

Por otro, los situados en la parte superior de la escala, es decir, los que corresponden a mayor elevación. Esto no da lugar a ambigüedad, ya que el primer caso siempre aparecerá cerca de tonos azules, mientras que el segundo se situará cerca de los marrones. No puede ser de otro modo, ya que equivaldría a que las curvas de nivel pudieran cortarse entre sí, lo cual sabemos que no es posible. El contexto de los colores circundantes sirve para eliminar la ambigüedad. En el mapa de coropletas, al no suceder necesariamente así, la ambigüedad permanecería y haría imposible discernir el significado de la simbología. En el caso de las isocoropletas, en la que la contigüidad espacial sí implica también contigüidad de clases, sí pueden utilizarse este tipo de esquemas, como ya vimos en la figura 28.12.

Por todo lo anterior, el uso de la componente valor es preferible frente al uso del tono a la hora de crear un mapa de coropletas para representar información cuantitativa.

28.6.5. Otros tipos de mapas

Existen muchos otros tipos de mapas, adecuados para representar tipos particulares de información. A pesar de su utilidad, son mucho menos frecuentes, especialmente dentro del ámbito SIG, ya que su implementación no es habitual y no resulta común crearlos con las herramientas usuales de estos. Algunos de estos tipos de mapas que resulta de interés reseñar son los siguientes:

- **Mapas dasimétricos**. Los mapas dasimétricos tratan de evitar las deficiencias de los mapas de coropletas, en los que los límites de las distintas áreas representadas no tienen relación con la variable con la que se trabaja, siendo limites arbitrarios tales como divisiones administrativas o territoriales. En los mapas dasimétricos las divisiones obedecen a la propia geografía de la variable. El principal inconveniente de estos mapas es el mayor esfuerzo que su preparación exige, así como el mayor conocimiento de la variable que resulta necesario para poder definir

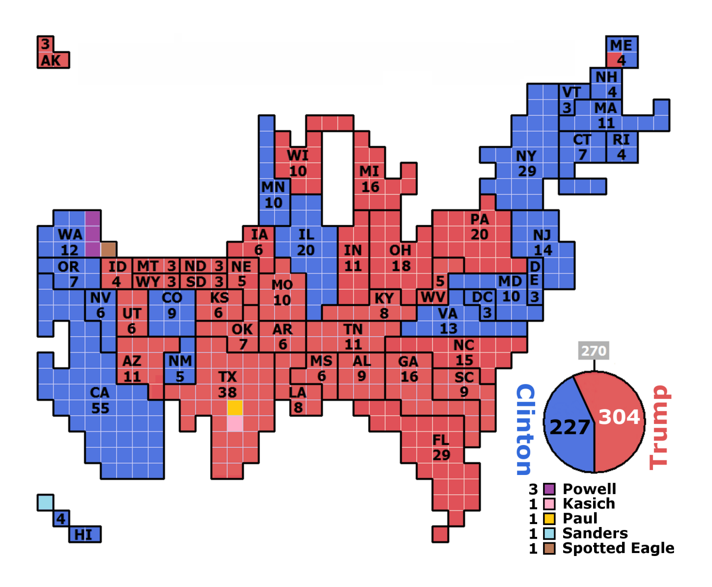


Figura 28.14: Un ejemplo de cartograma (Adaptado de Wikipedia).

las distintas zonas del mapa. Tradicionalmente se han empleado para representar la densidad de población, siendo poco usados para otras variables.

- **Mapas de flujo**. Los mapas de flujos representan movimientos de algún tipo de elemento, como por ejemplo las exportaciones de un producto o los desplazamientos de tropas en una campaña militar. El mapa de flujo aporta información sobre cómo se produce la distribución del elemento que se desplaza, la proporción o magnitud en que lo hace, así como también la ruta seguida, aunque este último factor no es habitualmente prioritario y suele representar más con carácter esquemático (indicando la relación entre los puntos de partida y destino del movimiento) que como verdadera información geográfica sobre el trayecto en cuestión. Algunos de los mejores ejemplos de mapas de flujo son los creados por Charles Joseph Minard (1781–1870), ingeniero francés pionero en su creación. Uno de esos mapas puede verse en la figura 28.13.
- **Cartogramas**. En los cartogramas, la información cualitativa se transmite mediante la modificación de las unidades de superficie, que se distorsionan para representar con su tamaño la magnitud de la variable en cuestión. Es decir, la variable visual tamaño se aplica directamente sobre las distintas unidades de superficie. En la figura 28.14 puede verse un ejemplo de cartograma en que los distintos estados de Estados Unidos se representan con un tamaño relativo a su número de votos electorales, evitando así que estados con poca densidad de población aparezcan en el mapa con una superficie y una preminencia que no les corresponde.

28.7. Resumen

Hemos visto en este capítulo cómo un mapa constituye una forma de comunicación visual, y cómo en esa comunicación existen una serie de factores a tener en cuenta para que la transmisión de la información entre emisor y receptor sea óptima. De especial relevancia en este sentido

es prestar atención a este último y tener siempre en cuenta el propósito del mapa que creamos.

Distinguimos dos tipos de cartografía: la cartografía de base y la temática. Esta última es la que crearemos con más frecuencia en un SIG. Las formas de cartografía temática están muy relacionadas con las características de la variable. Para el caso de variables cuantitativas, es importante agrupar adecuadamente los distintos valores en clases. Existen diversas formas de delimitar los intervalos correspondientes, siendo las más habituales el uso de intervalos iguales, intervalos naturales o intervalos basados en la media y la desviación típica de los valores en cuestión.

Dentro de los tipos de mapas temáticos más importantes encontramos los mapas de puntos, de símbolos proporcionales, de isolíneas y de coropletas, cada uno de ellos con sus características particulares. Los mapas de isolíneas son especialmente indicados para la representación de variables continuas, mientras que por su parte las variables de tipo razones se representan de forma especialmente adecuada mediante los mapas de puntos.

A la hora de componer un mapa existen diversos elementos que deben añadirse para facilitar su interpretación. Además de conocer la función de cada uno, es importante saber cómo situar estos sobre el lienzo del mapa, aprovechando correctamente el espacio e integrándolos adecuadamente.

Capítulo 29

La visualización en términos SIG

29.1. Introducción

El SIG es nuestro útil para visualizar la información geográfica y, como hemos visto, un útil muy potente y con numerosas posibilidades. Conocemos ya sus capacidades y limitaciones, pero no sabemos todavía cómo debemos trabajar con estas a la hora de crear una representación visual y, sobre todo, desconocemos la forma en que las particularidades de la información geográfica dentro de un SIG afectan a su visualización.

El concepto de capa, que resulta vital para otras tareas tales como el análisis, va a tener de igual modo una influencia directa en la creación de representaciones a partir de los datos de que disponemos, ya que la interpretación de estos datos está condicionada inevitablemente al modelo de datos empleado. Por ello, veremos en este capítulo los conceptos que ya conocemos de otros anteriores, pero aplicados al caso particular de aplicarlos dentro de un SIG, empleando las herramientas que este habitualmente incluye para esa tarea.

Puesto que conocemos ya un buen conjunto de operaciones sobre los datos espaciales, y estas operaciones forman parte integrante del SIG al igual que la visualización, estudiaremos asimismo cómo aprovechar algunas de estas operaciones de cara a la visualización de información geográfica. Es aquí donde reside una de las grandes virtudes del SIG, en que sus distintas capacidades están conectadas y son accesibles desde un mismo entorno. Hacer más patente esa relación entre ellas y ampliar así las posibilidades que un mismo juego de datos ofrece para ser representado es uno de los objetivos de este capítulo.

En conjunto, las capacidades que ofrece un SIG, incluso si en muchos casos no alcanzan la funcionalidad necesaria para satisfacer al cartógrafo profesional, permiten expandir las posibilidades de representación de una capa cualquiera y obtener visualizaciones distintas a las que son habituales en un mapa clásico. Del mismo modo que una aplicación CAD amplía las posibilidades del arquitecto o ingeniero a la hora de crear planos o realizar diseños industriales, o un programa de animación abre nuevos horizontes para un dibujante, los SIG expanden las posibilidad de creación y uso de cartografía para todo aquel que requiera visualizar la información geográfica.

Cuando la información geográfica se presenta en pantalla y dentro del contexto de la aplicación SIG correspondiente, lleva asociada además una interactividad de la que un mapa impreso carece, circunstancia que también debe tenerse en cuenta. Así, junto a las ideas básicas que ya hemos desarrollado en otros puntos dentro de esta parte del libro, es necesario añadir algunas adicionales para cubrir toda la gama de posibles resultados visuales que ahora encontramos gracias a los SIG.

Además de esa representación habitual de capas que constituye la adaptación de la cartografía clásica al entorno de un SIG, otras formas de visualización que encontramos en estos son novedosas y no tienen un equivalente en aquella. Aunque su relevancia es variable, algunas de ellas representan una parte importante de las funcionalidades actuales de los SIG, por lo que las trataremos por separado para ver cómo complementan a las herramientas de visualización habituales y cómo se integran con estas.

29.2. Visualización de capas vectoriales

La visualización de capas vectoriales en el seno de un SIG es similar a la labor de la cartografía clásica, en cuanto a que los objetos que se representan son del mismo tipo, esto es, objetos geométricos en forma de puntos, líneas y polígonos. A diferencia de las capas ráster, que no tienen un equivalente en un mapa clásico (no es probable que hayas visto un mapa previo a la aparición de los SIG con un aspecto como las imágenes de, por ejemplo, la figura 29.3), las capas vectoriales guardan mucha similitud con los elementos que un cartógrafo clásico plasma en un mapa. Las geometrías de las capas vectoriales son los objetos básicos sobre los que el cartógrafo aplica las variables visuales según vimos en el capítulo 27, y por tanto la manera de proceder es similar. Las herramientas que el SIG proporciona son aquellas que permiten modificar las variables visuales en función de las características asociadas a cada geometría a representar.

Un papel destacado en la visualización lo juega la tabla de atributos, ya que es la que contiene esas características que son necesarias para saber *cómo* representar cada objeto. El SIG provee la conexión entre los valores de los atributos y la representación visual, de forma que se interpretan aquellos para poder obtener los distintos tipos de mapas que ya conocemos del capítulo 28.

Al igual que el tipo de información es importante para escoger la clase de mapa a crear o la variable visual a emplear para la representación, el tipo de datos ha de estar correctamente definido en la tabla de atributos para poder emplearse como tal. Es decir, ha de ser coherente con la información que recoge.

Los atributos pueden contener en ocasiones no un valor que al interpretarse se convierta en una cualidad dada de una variable visual, sino esa cualidad directamente. En el caso de capas procedentes de aplicaciones CAD, es habitual que estas contengan algún campo con el color que ha de emplearse para representar esa capa, que puede venir indicado como un código que se deberá transformar después en el color correspondiente, o bien expresado directamente como el valor RGB de dicho color.

Respecto a las geometrias, es interesante hacer ver que, aunque los objetos geométricos que se representan son del mismo tipo que los objetos con los que trabaja el cartógrafo clásico, no ha de existir siempre obligatoriamente una identificación directa entre ambos. En otras palabras, el cartógrafo puede pintar un conjunto de líneas no mediante una línea sino como un conjunto de polígonos entre líneas, tal y como sucede en el mapa de isolíneas de la figura 28.12. Tanto el cartógrafo clásico como su equivalente moderno que emplea un SIG parten los dos de un conjunto de líneas, pero su forma de operar es distinta.

En un mapa clásico, se trazan las líneas y después se rellena con los colores correspondientes. En el SIG, no existe la posibilidad de «rellenar», ya que únicamente pueden aplicarse las variables visuales a las entidades, y el espacio intermedio entre líneas no es ninguna entidad que tengamos almacenada en la capa. No obstante, crear esa

representación resulta perfectamente posible en un SIG, aunque con un planteamiento distinto.

Puesto que queremos representar objetos de superficie, necesitamos una capa de polígonos. Obtener esta no debiera resultarnos complicado si conocemos las rutinas que vimos en la parte de procesos, que permiten convertir capas vectoriales en capas ráster, y viceversa. Podríamos, por ejemplo, rasterizar la capa de líneas, aplicar una reclasificación de sus valores para que queden en clases de la misma amplitud que la equidistancia de las isolíneas, y después vectorizar esas clases para obtener los polígonos que ya podríamos representar y colorear adecuadamente. Más adecuado es, no obstante, representar directamente la capa ráster así obtenida, sin necesidad siquiera de reclasificarla, ya que las clases las aplicaríamos en la visualización directamente. Esa es la metodología empleada para obtener la representación de la figura 28.12.

Conviven en la representación tanto una capa ráster como una vectorial. Lo relevante de este hecho es darse cuenta de las posibilidades que el SIG nos ofrece con sus funciones de análisis y mediante los procesos que hemos visto en una parte anterior del libro, y que pueden emplearse de igual modo para elaborar representaciones distintas. Ello aporta al SIG una flexibilidad que ha de aprovecharse cuando las capacidades de representación puramente dichas no nos ofrezcan la funcionalidad que necesitamos.

29.2.1. Etiquetado

El etiquetado representa una de las tareas más complejas a la hora de crear un mapa, ya sea con la ayuda de un SIG o sin ella, estimándose que puede llevar aproximadamente un 50 % del tiempo total de creación de un documento cartográfico [333]. La experiencia del buen cartógrafo queda patente en su trabajo con las etiquetas, pues es en esta labor donde más necesaria se demuestra, y en la que el criterio personal cobra una mayor importancia.

Siendo así, es lógico pensar que este es asimismo uno de los procesos en los que más difícil resulta proveer una solución automatizada, ya que trasladar al ordenador ese buen hacer del cartógrafo profesional no es en absoluto sencillo. Por ello, aunque un SIG puede incorporar herramientas para ayudar en el etiquetado, una gran parte de este trabajo sigue siendo necesario realizarla manualmente, y es por esta razón que conocer algunas ideas básicas al respecto es básico si queremos elaborar cartografía de una cierta calidad, ya que el SIG por sí mismo no va a poder llevar a cabo esta tarea de forma automática.

La premisa fundamental del etiquetado es situar las etiquetas de tal modo que estas no se solapen y que sea inmediato asociar su nombre al objeto geográfico que designan, así como a la importancia y propiedades de este. Para ello, necesitamos tres tipos de información a extraer de esos objetos:

- Dónde situar la etiqueta.
- Qué poner en la etiqueta.
- Cómo ponerlo.

Tratándose de capas vectoriales, toda esta información la extraeremos tanto de la geometría como de la tabla de atributos asociada. La más obvia es la relativa a qué debe ponerse en la etiqueta, que simplemente se tomará de alguno de los campos de la tabla que contenga los nombres de los distintos objetos.

Respecto a la posición, esta vendrá definida por la geometría y su georreferenciación, aunque solo parcialmente. La geometría nos da una indicación de la zona aproximada en la que debe situarse la etiqueta, ya que obviamente esta debe encontrarse cerca del objeto al que hace referencia, pero no constituye una información

suficiente, al menos para obtener un etiquetado óptimo más allá de la configuración más trivial.

Por ejemplo, en el caso de puntos cercanos, situar la etiqueta de estos centrada exactamente en cada uno de ellos hará que se solapen y oculten además a los propios puntos. Es necesario colocarlas cada una de ellas alejadas de los puntos en direcciones contrarias, para que no interfieran entre sí. La localización por tanto, no depende únicamente de las coordenadas del objeto, sino también de las de los objetos circundantes.

Buscar una disposición que evite estos solapes es una tarea en apariencia simple, pero compleja desde el punto de su implementación. Aun así, está presente en los SIG en mayor o menor medida, y en el caso de puntos, los resultados que se obtienen son de una calidad aceptable. El paso a otro tipo de geometrías, donde es necesario considerar otra serie de parámetros, hace aparecer unas circunstancias más difíciles de tratar, y la labor directa del cartógrafo es mucho más necesaria.

En el caso de capas de líneas, la posición de las etiquetas debe seguir el trazado de las líneas y su orientación, existiendo, no obstante, diversas opciones en lo que respecta a la posición con respecto a la propia línea. La línea ya no es un objeto puntual, y no existe por tanto una coordenada única que utilizar. El punto medio de la línea es la solución más inmediata como punto de referencia, pero no necesariamente la mejor. Pueden existir otras zonas a lo largo de la línea que resulten más relevantes y en las que sea más adecuado situar la etiqueta. En el caso de líneas muy largas, es conveniente repetir el nombre varias veces a lo largo de esta, para que no sea necesario seguirla hasta encontrar su nombre.

En el caso de líneas que se entrecruzan (calles, ríos, etc.), es importante evitar las ambigüedades. No es conveniente etiquetar una línea siempre que exista un cruce, pero un emplazamiento adecuado puede resultar suficiente para aclarar a qué línea hace referencia una etiqueta. Esto puede verse en la figura 29.1. En ambos casos, la etiqueta hace referencia al cauce que procede de la parte superior, que es el principal de los dos que confluyen, y por tanto también el que da nombre al segmento posterior al cruce. En el caso a), la mayor similitud en las direcciones puede inducir sin embargo a pensar que el nombre hace referencia al cauce después de la intersección y al segmento horizontal antes de esta. Si el etiquetado de este segmento horizontal, que es un cauce de nombre distinto, no se encuentra suficientemente cerca del cruce, puede entonces pensarse que la etiqueta hace referencia a él también lo cual no es adecuado. Un emplazamiento tal como el mostrado en el caso b) aclara esta situación de forma elegante.

Para el caso particular de las isolíneas, se recomienda situar la etiqueta sobre la propia línea, ya que facilita su lectura, especialmente en el caso de que aparezcan varias isolíneas separadas por poca distancia, como puede verse en la figura 29.2. Además, deben situarse las etiquetas de las isolíneas contiguas de tal forma que puedan leerse conjuntamente, para que sea sencillo interpretarlas en conjunto y apreciar sin dificultad la equidistancia y la dirección en la que los valores aumentan o disminuyen.

Si la línea presenta cambios de dirección bruscos, es difícil hacer que la etiqueta siga la línea sin tener un aspecto «roto». Suavizar las líneas es una opción en este caso, al menos para usarlas como líneas base sobre las que situar las etiquetas.

En el caso de polígonos, existe igualmente el problema de seleccionar un punto para emplazar la etiqueta. El centroide del polígono es la opción más inmediata, aunque no necesariamente ha de caer dentro de este si se trata de un polígono cón-

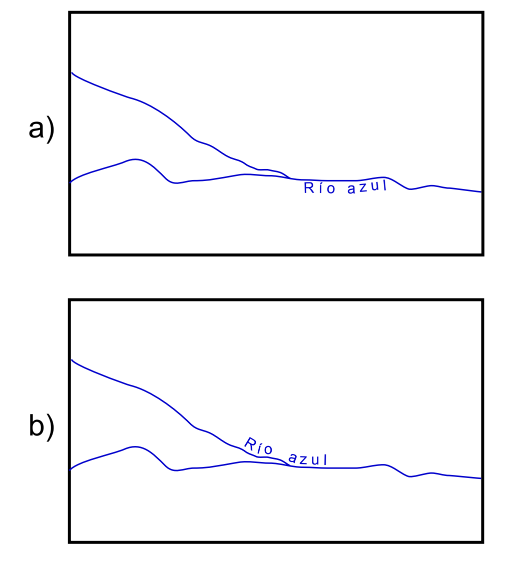


Figura 29.1: La posición de la etiqueta sobre una línea en un cruce puede dar lugar a ambigüedades (a) o a situaciones bien definidas (b).

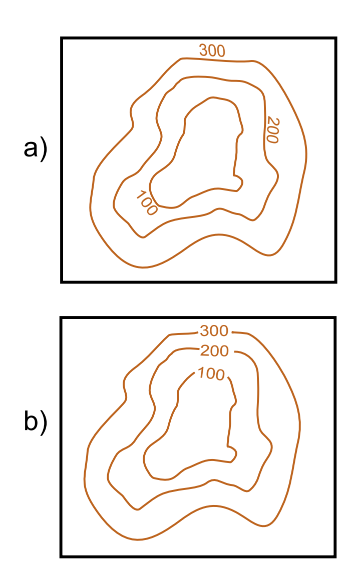
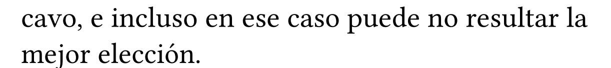


Figura 29.2: Etiquetado de isolíneas. Deben situarse las etiquetas cercanas entre sí y sobre la línea, como en el ejemplo b)


cavo, e incluso en ese caso puede no resultar la mejor elección.

Como puede verse, para tomar este tipo de decisiones es necesario tener en cuenta no solo la posición del objeto y la de los circundantes, sino también «entender» qué es lo que estamos representando y qué otra información tenemos alrededor, lo cual resulta más complejo de trasladar a la aplicación SIG para que pueda hacerlo de forma automática. Tal y como comentábamos, la intervención del cartógrafo es en este caso imprescindible para incorporar este tipo de circunstancias y aportar al mapa la calidad que un mecanismo automático de etiquetado no es capaz por el momento de ofrecer.

Una vez se ha definido la posición más adecuada para las etiquetas, es necesario decidir cómo representar cada una de ellas. Algunas etiquetas son más relevantes que otras, y la claridad con la que una etiqueta transmite su información depende en gran medida de cómo esta se escribe. Los conceptos de la tipografía son de relevancia en este caso, y son a los que debemos acudir. He aquí algunos de ellos.

- El uso de las variables visuales que conocemos es limitado en el caso de las etiquetas y, salvo el tamaño, no suelen emplearse para diferenciar unas de otras o darles más importancia.
- El uso del tono o el valor debe llevarse a cabo con precaución. La legibilidad de la etiqueta, no obstante, está en relación con el fondo, ya que el color de este puede dificultar su lectura, y en esta situación es a veces necesario usar uno u otro tono para garantizar esa legibilidad. La etiqueta siempre está en un primer plano, por lo que el resto del mapa bajo ella y en su entorno forma parte del fondo. Como ya vimos, un adecuado contraste entre fondo y figura es importante, por lo que variar el color de una etiqueta puede a veces ser necesario para que esta pueda leerse correctamente.

Algunos elementos se etiquetan sistemáticamente con colores establecidos, como en el caso de los ríos, del mismo color azul que la propia geometría de estos.

- El tamaño es la forma principal de jerarquizar las etiquetas y darles más importancia. Se puede aplicar directamente sobre el tamaño de la fuente, aunque también es posible hacerlo sobre el grosor (fuente normal o negrita).
- El uso de mayúsculas o minúsculas sirve igualmente para conceder más importancia a unas u otras etiquetas.
- La separación entre caracteres se puede modificar para hacer que la etiqueta cubra un espacio mayor a lo largo de un objeto lineal, eliminando en ocasiones la necesidad de un etiquetado múltiple de esta. Un espaciado mayor también aporta mayor énfasis a la etiqueta. También se puede optar por un espaciado menor en etiquetas menos importantes, o en zonas con alta densidad, para disminuir el espacio que estas ocupan y evitar solapes.
- El uso de fuentes artísticas o decorativas no es recomendable. Se deben utilizar fuentes sencillas y que sean lo más legibles posible.

La información necesaria para realizar todos estos ajustes a las etiquetas debe estar contenida en la tabla de atributos de la capa. Así, podemos incluir en esta campos que indiquen el ángulo en el que se escribe la etiqueta, el tamaño a utilizar o la separación de carácter, entre otras características. Incluso la propia posición puede especificarse de esta forma. En caso de existir estos valores, el SIG los usará en lugar de aquellos que resultarían de la aplicación de los algoritmos de etiquetado automático de que disponga, entendiendo que el ajuste manual es de mejor calidad. Dado que este tipo de configuración es habitual si se desea crear un mapa de calidad, algunos SIG permiten la incorporación de capas de etiquetado, que contienen toda la información necesaria para el establecimiento de etiquetas, de forma que estas se incorporan al mapa por separado y no a partir de los objetos a los que hacen referencia y sus atributos. Esta manera de proceder, no obstante, es más laboriosa.

En resumen, la tarea de etiquetar un mapa es compleja y normalmente va a requerir una cierta cantidad de trabajo manual por parte del creador del mapa. Los SIG disponen de herramientas para automatizar una parte de este trabajo, aunque la implementación de estas herramientas es muy variada, y encontramos desde aplicaciones con poco más que un sistema trivial de etiquetado hasta sistemas complejos altamente configurables. En cualquier caso, incluso en el más avanzado de los programas, es muy probable que debamos llevar a cabo algún tipo de modificación o que debamos especificar manualmente algunos de los parámetros que el SIG emplea para llevar a cabo un etiquetado automático o semi-automático.

29.3. Visualización de capas ráster

Las capas ráster son, en lo que a visualización respecta, las que resultan más novedosas si las comparamos con lo que encontramos en un mapa clásico. A diferencia de las capas vectoriales, compuestas por elementos que sí aparecen en estos mapas y cuya estructura lógica se asemeja mucho a la estructura gráfica de un mapa a base de símbolos puntuales, lineales y de superficie, las capas ráster dan lugar a representaciones que no resulta frecuente ver en la cartografía clásica.

La cartografía clásica, especialmente la relativa a lo que denominábamos cartografía base, se encarga de recoger qué es lo que hay en una determinada porción de terreno, llevando esto a cabo mediante la representación de una serie de objetos que se corresponden con aquello que encontramos en ese terreno. Este es un enfoque

mucho más acorde con el modelo de representación vectorial, y más alejado del modelo ráster. La representación gráfica de variables continuas, las cuales se aprovechan plenamente con el modelo ráster, no es objeto tradicionalmente de la cartografía, y de serlo se representan mediante geometrías simples, tales como las líneas en un mapa de isolíneas. Es decir, para el cartógrafo clásico, e independientemente del tipo de variable a representar, los datos se manejan en un modelo de representación de tipo vectorial.

Esto obedece principalmente al gran detalle que tiene una capa ráster, el cual hace inviable el uso de un planteamiento similar a la hora de crear un mapa sin la ayuda de un SIG. El cartógrafo puede trazar unas isolíneas sin dificultad para representar la topografía, pero dividir el lienzo del mapa en miles de pequeños cuadrados y colorear cada uno de un color según su elevación es por completo inviable. Más aún, disponer de los datos a representar en este caso (que constituirían un MDE), resulta también imposible si no se dispone de un ordenador para calcularlo.

Por todo ello, las capas ráster nos van a permitir crear representaciones algo distintas a las habituales en la cartografía clásica y, aunque las diferencias conceptuales con respecto a la visualización de capas vectoriales son pocas, hay algunas ideas que deben detallarse.

Formalmente, y al igual que de cara a su análisis, podemos considerar una capa ráster como una capa vectorial de polígonos (cuadrados en este caso). No obstante, e igual también que para el análisis, la regularidad de la capa ráster es el elemento clave que aporta la diferencia más importante, y en el que reside la particularidad de ese modelo de representación.

Si a cada uno de los polígonos cuadrados de los que se compone una capa ráster le asignamos un color, podemos considerar que el mapa resultante es equivalente a un mapa de coropletas, aunque con tres características peculiares: las unidades tienen un mismo tamaño todas ellas, este tamaño es normalmente muy pequeño y tiene dimensiones muy reducidas en la representación, y las unidades están situadas de forma regular en una malla. Estas características hacen que algunos de los inconvenientes de los mapas de coropletas no se presenten, y permiten un uso distinto de las variables visuales.

Por ejemplo, el uso del tono como variable ordenada, que ya vimos que en ciertos casos sí resulta adecuado, se puede dar en las capas ráster. Como ya se mencionó al desarrollar las variables visuales, puedes encontrar abundantes ejemplos de representaciones así en los capítulos de la parte dedicada a procesos dentro de este libro. La regularidad de la malla de celdas, junto con la autocorrelación espacial y la continuidad de una variable a representar, hacen que cada celda esté rodeada de otras de valores similares, lo que aporta también una continuidad visual que puede aprovecharse para emplear esquemas ordenados basados en el tono.

Tanto si se usa el valor como si se usa el tono, otra de las consecuencias de la estructura de una capa ráster, y en particular del pequeño tamaño de sus celdas es el hecho de que resulta de interés aumentar el número de clases en que dividimos los valores de la variable para asignarles el correspondiente valor o tono como variable visual. La mayor resolución espacial con la que trabajamos se puede acompañar también de una mayor resolución cromática para obtener representaciones de mayor riqueza.

Mencionábamos en el apartado 28.4.1 que no se recomienda un número de clases mayor de 7 u 8, ya que haría complejo el identificar cada una de ellas en la leyenda y conocer la cantidad exacta que se representa. Ello no significa, sin embargo, que el ojo humano no pueda distinguir más de 8 valores distintos de un tono dado. Si

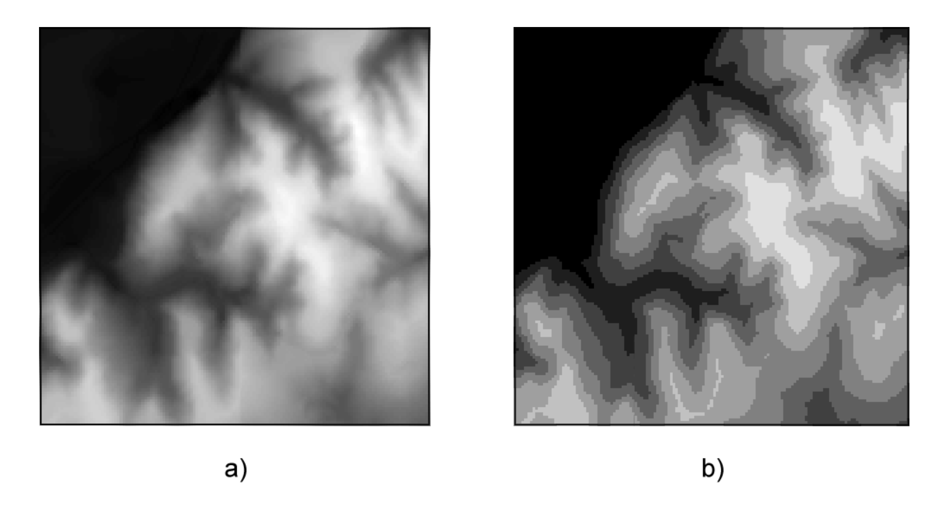


Figura 29.3: Representación de una capas ráster mediante 255 (a) y 8 (b) clases.

extendemos el número de clases, podemos lograr un efecto de transición suave entre los colores de las distintas celdas y eso, aunque no facilite la identificación de un color concreto con su valor asociado de la variable representado, crea una representación mucho más informativa. Puede verse esto claramente en la figura 29.3

La representación de la parte izquierda de la figura, con un total de 255 clases, da más detalle sobre la distribución de la variable a lo largo del mapa que la de la parte derecha, con 8 clases. Saber en qué rango de valores se encuentra una zona dada del mapa puede resultar más difícil e impreciso, pero a cambio tenemos más detalle. En un mapa de coropletas, con unidades grandes bien diferenciadas, usar más clases no aporta más detalle sobre la distribución de la variable, ya que falta esa suavidad en las transiciones entre unidades. En una capa ráster, por el contrario, la ganancia es notable.

29.4. Combinación de capas

Argumentábamos en los primeros capítulos de este libro, cuando presentábamos el concepto de *capa*, que el verdadero éxito de este concepto es la separación de los distintos tipos de información geográfica, atomizando esta en unidades autocontenidas que guardan tan solo la información relativa a una variable o fenómeno concreto. Así, cuando adquirimos un mapa impreso, obtenemos muchas variables distintas que no podemos separar, pero en un SIG, y con la información ya separada en esas capas, la situación es muy distinta, dando lugar a un manejo más estructurado y eficaz.

Pese a esto, resulta claro que a la hora de representar la información geográfica, una capa aislada no constituye la forma óptima de visualizar esta. Si en un mapa encontramos elementos variados, ello no obedece a la mera economía de espacio, sino a que añadir información adicional a la de esa capa que queremos representar nos ayuda a entenderla mejor. Los procesos que tienen lugar en el espacio están relacionados unos con otros, y visualizar esas relaciones aporta una mayor riqueza a la visualización, haciendo que sea más sencillo extraer la información contenida en una capa. Podemos ver un claro ejemplo de esto en la figura 29.4

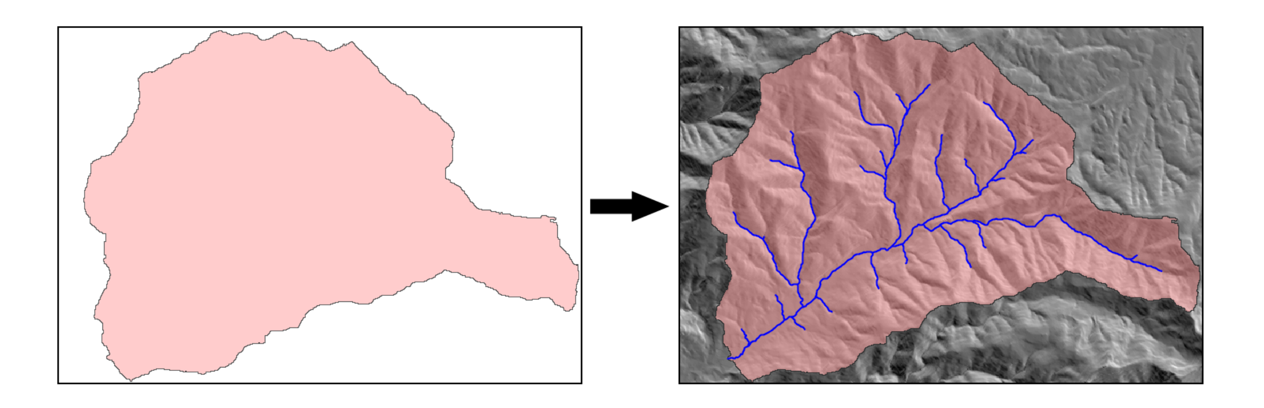


Figura 29.4: Añadir capas adicionales que complementen a aquella que resulta de interés nos ayuda a interpretar mejor esta y a lograr una representación más eficaz.

La capa que representa la cuenca vertiente a un punto, y que contiene un solo polígono, resulta mucho más útil visualmente si la acompañamos de elementos básicos como el relieve y los cauces principales. La imagen de la derecha es autoexplicativa y se ve claramente gracias al relieve que el polígono delimita la cuenca. En la de la izquierda esa información no puede deducirse únicamente de la capa de interés.

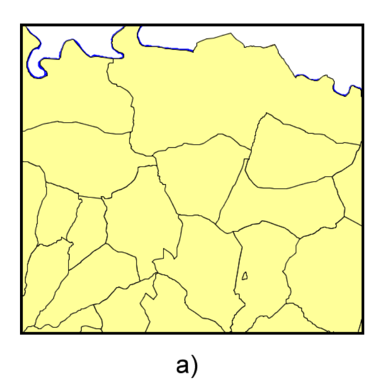
Aunque sencillo de llevar a cabo en lo que a manejo del SIG respecta, combinar capas es un proceso que también debe realizarse con conocimiento, y en el que, si se realiza correctamente, las diferencias pueden ser notables. No solo se trata de dar espacio dentro del mapa a toda la información que esas capas contienen, sino que exista una sinergia entre ellas en la medida de lo posible, para que se complementen mutuamente como partes de un conjunto. Veremos en este apartado algunas ideas a tener en cuenta en este sentido.

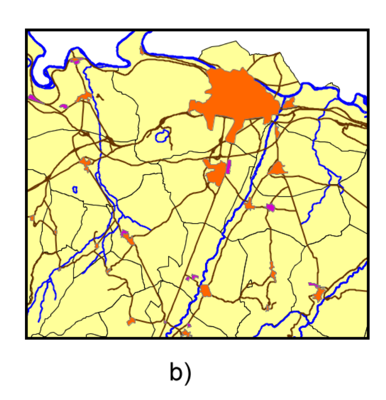
El primer aspecto a considerar es el orden de las capas, que indica cómo se disponen estas las unas sobre las otras y definen el orden de pintado. Si una misma zona está ocupada por elementos de varias capas, solo serán visibles los correspondientes a la capa superior, ya que la representación de los pertenecientes a las demás quedará oculta. El efecto es el mismo que si pintáramos en un papel algo y encima de ello pintáramos después algo distinto. Tan solo veríamos esto último.

La figura 29.5 muestra un claro ejemplo de lo anterior en el que se puede apreciar la diferencia que supone variar el orden de las capas.

A pesar de estar construidas a partir de las mismas capas, las representaciones mostradas en la figura son muy distintas como documentos cartográficos y no proporcionan la misma información. Así, en el caso b), prácticamente toda la información esta «oculta», ya que hay una capa que la cubre. En el caso c) sucede que las zonas urbanas (en marrón) están situadas por encima de las capas de ríos y vías de comunicación, dando la sensación de que estas últimas desaparecen al entrar en dichas áreas urbanas. Este puede ser un efecto deseado en ciertos casos, para enfatizar las zonas urbanas y su contorno, pero la representación es menos informativa en caso de que quiera detallarse el trazado de cauces y carreteras.

Se ve claramente que el orden de pintado es importante, y que un correcto orden es vital para acomodar todos los elementos a representar y





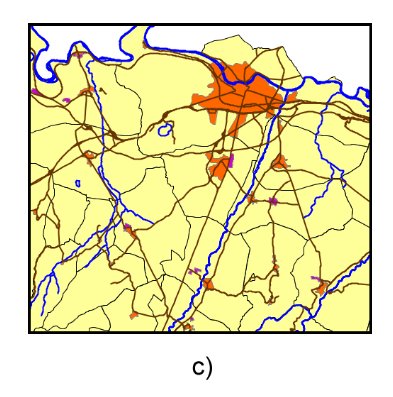


Figura 29.5: Variar el orden de las capas puede suponer un cambio radical en la representación final obtenida. Todas las imágenes proceden del mismo conjunto de capas, pero las representaciones son muy distintas.

que cada uno cumpla su labor como elemento informativo.

Sabemos que las capas ráster llenan todo el espacio y contienen valores en todas sus celdas (o píxeles en el caso de imágenes). Por ello, van a tapar lo que se sitúe por debajo de ellas y no resulta buena idea situarlas en lo alto del orden de pintado. En su lugar, se deben considerar como capas base sobre las que situar las restantes, de tal modo que no impidan a estas visualizarse correctamente.

Con un razonamiento similar, podemos establecer la mejor forma de ordenar las capas vectoriales, situando por norma general los polígonos en primer lugar, y encima de estos las líneas y los puntos respectivamente. Esta regla es, lógicamente, muy genérica, y en cada situación se ha de evaluar la conveniencia de adoptar otra disposición, siempre con objeto de evitar que unas capas dificulten la correcta interpretación de otras.

En ocasiones, un determinado orden viene impuesto por el significado que tienen las capas. Por ejemplo, si nuestro mapa contiene una capa con la red de drenaje y otra con carreteras, lo lógico y habitual es que las carreteras estén por encima de los ríos, ya que lo normal es que pasen por encima de estos y no al contrario. En la practica cartográfica, este tipo de situación se resuelve simbolizando de forma particular este tipo de coincidencias, como se muestra en la figura 29.6. Esto requiere en el SIG unas capacidades avanzadas de edición gráfica. Algunos SIG incluyen no solo esas capacidades, sino también funcionalidades que crean automáticamente esos elementos gráficos en función del análisis de las capas, de tal modo que automatizan la tarea.

Una funcionalidad de que disponen los SIG para la combinación de capas es el uso de transparencias y semi-transparencias. Estas se pueden aplicar tanto a capas ráster como vectoriales, de forma que puede verse a través de ellas y así presentar la información de otras capas que se encuentren por debajo. Por ejemplo, la representación mostrada en la figura 29.4 hace uso de esta técnica. El polígono que delimita la cuenca vertiente es semi-transparente, de tal modo que la capa de relieve sombreado que está debajo puede verse, dando la sensación de que sigue ese relieve.

Si se usa semi-transparencia para una capa temática (por ejemplo, en un mapa de coropletas), no debe perderse de vista que el color varía respecto al original que ha sido asignado a cada

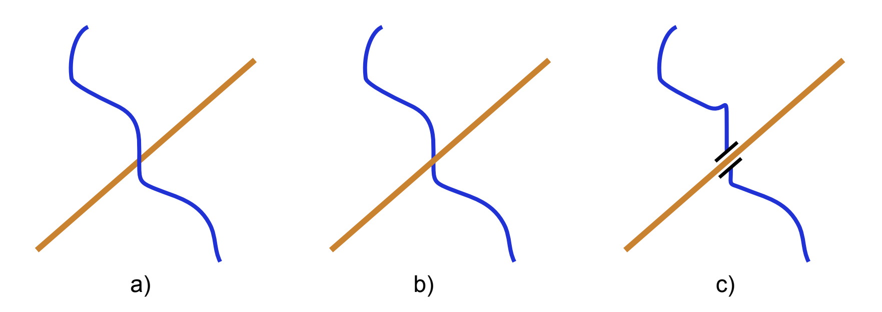


Figura 29.6: Representación errónea (a) y correcta (b) de capas al combinar una de carreteras (en marrón) y otra de red de drenaje (en azul). La inclusión de un elemento que simbolice el cruce (c) supone, no obstante, una mejor solución desde el punto de vista cartográfico.

polígono, ya que se «mezcla» con el color de cada píxel correspondiente a la representación de las capas inferiores. Esto puede resultar confuso a la hora de interpretar las componentes del color, ya que no coincidirán con las mostrada en la leyenda. Más aún, y como puede también apreciarse en la figura 29.4, el color del polígono, que debería ser único, no lo es, ya que la parte que se transparenta a través de este no es uniforme. En el caso mostrado en la figura, este hecho no tiene importancia, pero debe considerarse al representar otro tipo de variables en las que el color tiene un significado definido, para garantizar que ese significado se transmite de igual modo.

En el caso de una capa ráster, puede aplicarse una transparencia total, haciendo que determinadas partes de esta no se representen. A pesar de que la capa ráster contiene información en todas las celdas de su extensión, no todas se representan. Esto es especialmente útil para capas de tipo categórico. La figura 20.10 es un buen ejemplo de esto. En ella, la capa que contiene el área de influencia es una capa ráster, ya que ha sido creada mediante un análisis ráster (repasa el apartado para ver cómo se ha calculado, si no lo recuerdas). Sin embargo, se puede combinar con la capa de pendientes, ya que solo se pintan las celdas correspondientes a dicho área de influencia pero no las restantes. Para llevar esto a cabo se suele asignar la transparencia a un valor o serie de valores definidos, habitualmente al valor que codifica la ausencia de datos (que en este caso es el empleado para codificar aquellas celdas que no forman parte del área de influencia calculada).

El uso de transparencia sirve también para combinar imágenes que se solapan y eliminar las partes de estas que no contienen información. Como vimos en la sección 5.3.1, la forma de la imagen es siempre rectangular y tiene una orientación fija. Esto no ha de coincidir obligatoriamente con la información que contiene, siendo necesario en ese caso rellenar las áreas sin información con algún tipo de valor. A la hora de combinar capas, esos valores de relleno no interesa representarlos.

La división horizontal de los datos puede dar lugar a problemas en el caso de capas vectoriales o capas ráster distintas de imágenes, cuando sea necesario establecer unas características de representación en función de sus atributos, en caso de que la información acerca de una variable se encuentre dividida en varias capas. Un SIG incorpora habitualmente herramientas para que estas capas, así divididas para una mejor gestión, puedan unirse en una única, y al hacer esto, la capa resultante tendrá asignado un esquema de representación también único, de forma que toda

ella se visualizará de forma coherente. En tal caso, no encontramos problema alguno.

En el caso, sin embargo, de trabajar con las capas de forma independiente, y si estas han de combinarse en una misma representación, es necesario que los esquemas de representación sean coherentes unos con otros, para que en la representación global aparezcan como una única capa de información. De modo contrario, la representación será ambigua y confusa, y no mostrará de la forma adecuada la información que esas capas contienen. No considerar esta circunstancia lleva a errores tales como los mostrados en la figura 29.7.

En el caso de la derecha, dos MDE se representan con una misma gradación de colores. Se usa una representación por intervalos, pero, debido a que los valores extremos a partir de los cuales se crean dichos intervalos son distintos, estos intervalos resultan también distintos, y un mismo color representa un valor de elevación diferente en cada capa. Por esta razón, se hace muy patente la línea de unión entre ambas capas, ya que, pese a que existe una continuidad suave entre los valores, no lo es así en lo que respecta a su representación.

El mismo par de capas puede representarse de forma correcta sin más que establecer un único conjunto de intervalos para ambas, de tal modo que los valores máximos y mínimos entre los que se sitúen sean los máximos y mínimos absolutos del conjunto de capas.

En el caso de la izquierda (que ya se describió en el apartado 19.4 dedicado a la operación de juntar capas vectoriales) se presenta el mismo error, aunque no resulta tan patente a primera vista como en el anterior. La representación está realizada a partir de cinco capas de datos, una para cada continente, asignando colores en función de la población de cada país y con un total de 10 intervalos. Aunque la representación no revela ningún problema tal como la línea de sutura entre las capas ráster del ejemplo a), es incorrecta, ya que países con poblaciones muy distintas se representan con un mismo color. Así, Alemania, el país más poblado del contiene europeo, y China, el más poblado de Asia, tienen el mismo color a pesar de este último tiene más de quince veces más habitantes que el primero.

Una vez más, los intervalos empleados no son coherentes entre sí. En la representación de la derecha de la figura puede observarse el resultado tras haber ajustado convenientemente los parámetros de representación del conjunto de capas. Nótese que, pese a ser correcto desde este punto de vista, el mapa es poco informativo. La división en intervalos iguales que se ha empleado no resulta una buena opción en este caso debido a la presencia de unos pocos países con mucha más población que el resto. El uso de intervalos naturales o percentiles habría dado lugar a una representación más útil.

29.5. Particularidades de la representación en pantalla

Tanto para las representaciones en papel como para las representaciones en pantalla se siguen unos mismos principios a la hora de diseñarlas, pero estas últimas presentan algunas características particulares que hacen necesario tener en consideración otros factores. Esto es especialmente cierto cuando consideramos que esa representación en pantalla se realiza desde dentro de un SIG como parte de una sesión de trabajo con este (es decir, que no estamos, por ejemplo, visualizando un mapa escaneado con una aplicación de edición de imágenes sino trabajando en el SIG y creando la visualización como un elemento más de ese trabajo).

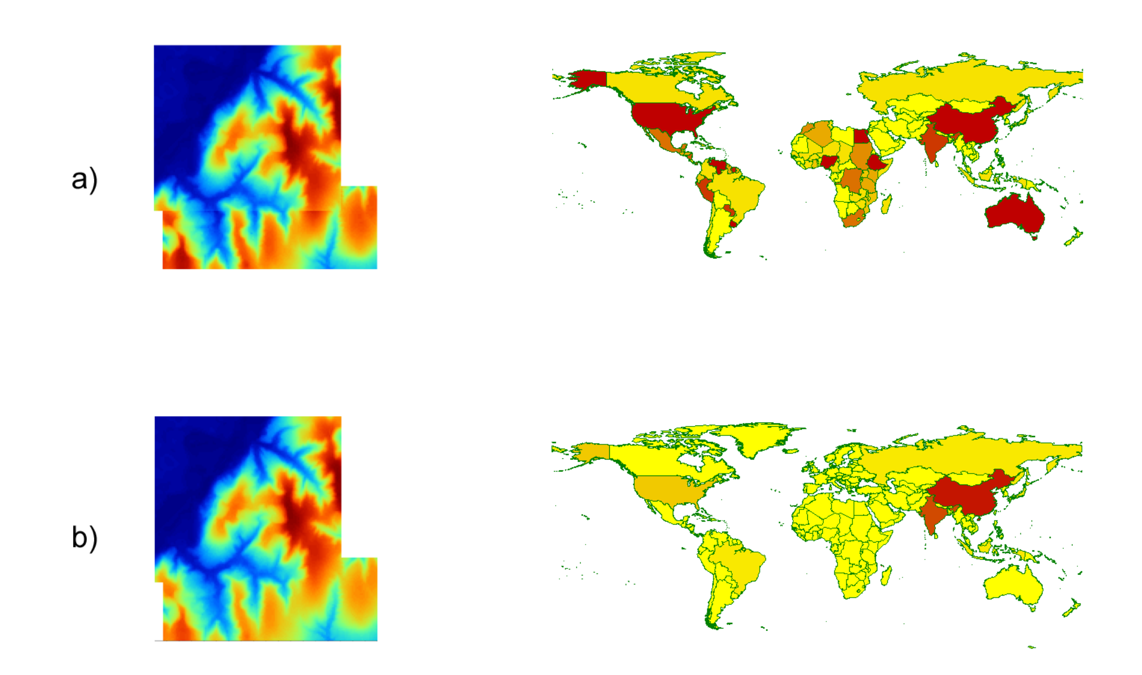


Figura 29.7: a) dos representaciones incorrectas de conjuntos de capas, debido a incoherencia entre los parámetros de representación empleados en cada una de ellas. b) representación correcta y homogénea con parámetros de representación comunes.

Podemos distinguir dos bloques fundamentales de diferencias que hacen que un mapa pensado para ser visualizado en la pantalla mientras ejecutamos un SIG no deba diseñarse exactamente igual que si estuviera pensado exclusivamente para ser utilizado en un soporte impreso: la baja resolución de la pantalla y la interactividad de la propia representación.

El primer aspecto a considerar es la baja resolución de una pantalla en comparación con un documento impreso. Mientras que sobre el papel un mapa puede imprimirse a una resolución de varios cientos de puntos por pulgada (dpi), en la pantalla la resolución viene limitada por el tamaño de los píxeles, que es mucho mayor que el del mínimo punto que se consigue imprimir por medios mecánicos. En una pantalla, la resolución es del orden de 100 píxeles por pulgada. Por eso, si te acercas a la pantalla de tu ordenador, puedes ver los píxeles individuales si tienes cierta agudeza visual. Por el contrario, incluso con una impresora de uso doméstico, distinguir el mínimo punto que esta es capaz de imprimir está por encima de la capacidad del ojo humano. Esto quiere decir que el papel permite una definición mucho mayor que la pantalla, ya que incluso los elementos de menor tamaño del mapa van a estar dibujados con una serie puntos de menor tamaño que permiten lograr una nitidez muy elevada.

A la hora de preparar cartografía impresa, la resolución no es un problema, ya que las capacidades de que se dispone superan a las necesidades que el cartógrafo puede tener. En la pantalla, sin embargo, algunos elementos pueden no aparecer con suficiente claridad y, aunque en papel cumplan su función correctamente, es conveniente sustituirlos por otros más adecuados cuando no se trabaja sobre un medio impreso. Los siguientes

son algunos de los elementos que deben evitarse o, al menos, emplearse de manera distinta a la hora de crear representaciones en pantalla.

- Fuentes con ornamentos tales como sombreados. Si son de pequeño tamaño, el sombreado no puede pintarse con suficiente nitidez y perjudica la legibilidad del texto.
- Fuentes con *serifas*. Las serifas (Figura 29.8) están pensadas para hacer más cómoda la lectura del texto impreso cuando este tiene una longitud considerable tal como en un libro, y consisten en pequeños adornos generalmente situados al final de las líneas. Por su pequeño tamaño, no se representan con suficiente definición en la pantalla, lo que causa pérdida de legibilidad. Por ello, se recomienda el uso de fuentes sin serifas en documentos pensados para visualizarse en pantalla, tales como paginas Web o como un mapa dentro de un SIG.
- Rellenos con tramas de paso muy fino. Si las líneas de un tramado están muy juntas, la baja resolución de la pantalla puede ser insuficiente para separarlas, haciendo difícil para el observador reconocerlas.
- Punteados. Al igual que en el caso anterior, si el punteado no tiene un paso suficiente, puede no resultar evidente la discontinuidad de la linea, creándose una representación ambigua.

El segundo aspecto a considerar es el relativo a la interactividad de las representaciones. A diferencia de un mapa impreso, en un SIG lo que vemos no es un elemento estático, sino dinámico. En este contexto, *dinámico* no quiere decir que el mapa cambie o que represente un proceso dinámico (que también es posible, como veremos más adelante en otro apartado de este capítulo), sino que el usuario puede alterarlo utilizando por lo menos las herramientas más fundamentales que proporcionan interactividad, tales como el desplazamiento, el acercamiento o el alejamiento, según ya vimos en el apartado 23.2.2. Este hecho hace que aparezcan algunos problemas, entre los que destacan los relacionados con el rendimiento y aquellos que derivan de la posibilidad de variar sensiblemente la escala de representación.

Respecto al rendimiento, no debe olvidarse que cada vez que formamos la imagen de un mapa en la pantalla (algo que sucede cada vez que ajustamos el encuadre mediante esas herramientas interactivas), el SIG ha de realizar un gran número de cálculos correspondientes a operaciones como las siguientes:

- Remuestreo de las imágenes.
- Asignación de colores o patrones a los distintos elementos (geometrías en capas vectoriales o celdas en capas ráster) en función de los valores asociados a estos.
- Dibujado de geometrías.

En función de la complejidad y el tamaño de las capas que estemos representando, así como del número de estas, generar esa representación puede suponer un volumen muy elevado de operaciones, lo cual hará poco fluido el trabajo en el SIG, llegando incluso a hacer inoperativa la propia interactividad del programa en un caso extremo. Cuando esto sucede, es necesario sacrificar algo de precisión y rigor cartográfico en beneficio del rendimiento, especialmente cuando la falta de rendimiento y la lentitud del sistema nos dificulten la realización de otras operaciones tales como, por ejemplo, el análisis de esas mismas capas que representamos, o incluso la propia navegación.

Trabajar con capas de menor detalle —por ejemplo, capas ráster de menor resolución o capas vectoriales con líneas simplificadas (véase 19.7)— es una solución a este problema, aunque no necesariamente excluye la posibilidad de trabajar con las capas originales. Un planteamiento multi-escalar en el que, según la escala, se visualicen unas u otras capas, es una solución frecuente a esta problemática. Vimos estas ideas en el

AaBbCc

Figura 29.8: Concepto de serifa.

apartado 3.5.2, donde presentamos el concepto de *pirámide* como recurso empleado en estos casos para el trabajo con capas ráster.

También se puede aumentar la velocidad de dibujado utilizando colores lisos en lugar de tramas, y evitando los textos de gran tamaño o los símbolos complejos que provengan de imágenes muy detalladas y de gran tamaño.

Por último, el hecho de que la escala de la representación pueda variar según la voluntad del usuario puede causar problemas con algunos de sus elementos tales como símbolos o etiquetas de texto. Si todos los elementos del mapa se escalan proporcionalmente, una reducción importante de escala disminuirá el tamaño del texto hasta hacerlo ilegible. Por el contrario, si aumentamos la escala el tamaño puede ser excesivo. La figura 29.9 muestra este hecho.

El mismo problema sucede en el caso de emplear símbolos. Si, por ejemplo, tenemos una capa de puntos con la localización de bocas de incendios y representamos cada uno con un pequeño dibujo de una de ellas, al aumentar el tamaño de cada icono se perderá definición, mientras que al disminuirlo la pantalla no tiene resolución suficiente para dibujarlo correctamente y no se identificará su forma. En general, el empleo de símbolos puntuales de este tipo se desaconseja a la hora de representar cartografía en pantalla.

Una solución a esto es especificar un tamaño absoluto de estos elementos que no varíe con la escala. Es decir, que un símbolo o una etiqueta de texto tengan siempre el mismo tamaño en pantalla y ocupen los mismos píxeles. A escalas bajas, sin embargo, este método puede dar lugar a representaciones saturadas, como se observa en la figura 29.10. Este problema es más notable si se tiene en cuenta que en pantalla se emplean generalmente tamaños de letra más grandes que en un mapa impreso, por lo que se debe reducir la cantidad de texto mostrado para evitar una densidad de etiquetas demasiado elevada.

Las particularidades que hemos visto en esta sección se refieren a la representación en la pantalla de un ordenador de sobremesa o portátil, pero, como vimos en el capítulo 25.1, los SIG sobre dispositivos móviles tienen a su vez sus propias características en lo que a dispositivos de representación respecta. Por ello, y según los casos, todo lo visto en este apartado debe considerarse de modo específico en estos casos, añadiendo los condicionantes que este hecho puede implicar en las distintas funciones de representación.

29.6. Visualización tridimensional

La visualización tridimensional es una de las tendencias más importantes dentro del ámbito SIG en la actualidad. Aunque el SIG de escritorio sigue siendo fundamentalmente una herramienta 2D, las aplicaciones con capacidades 3D van ganando relevancia al tiempo que incorporan cada vez más funcionalidades que las acercan a las del SIG de escritorio completo.

Además de su mejor capacidad para incorporar de forma realista los elementos geográficos (que son tridimensionales, así como los fenómenos que los originan), una de las razones induda-

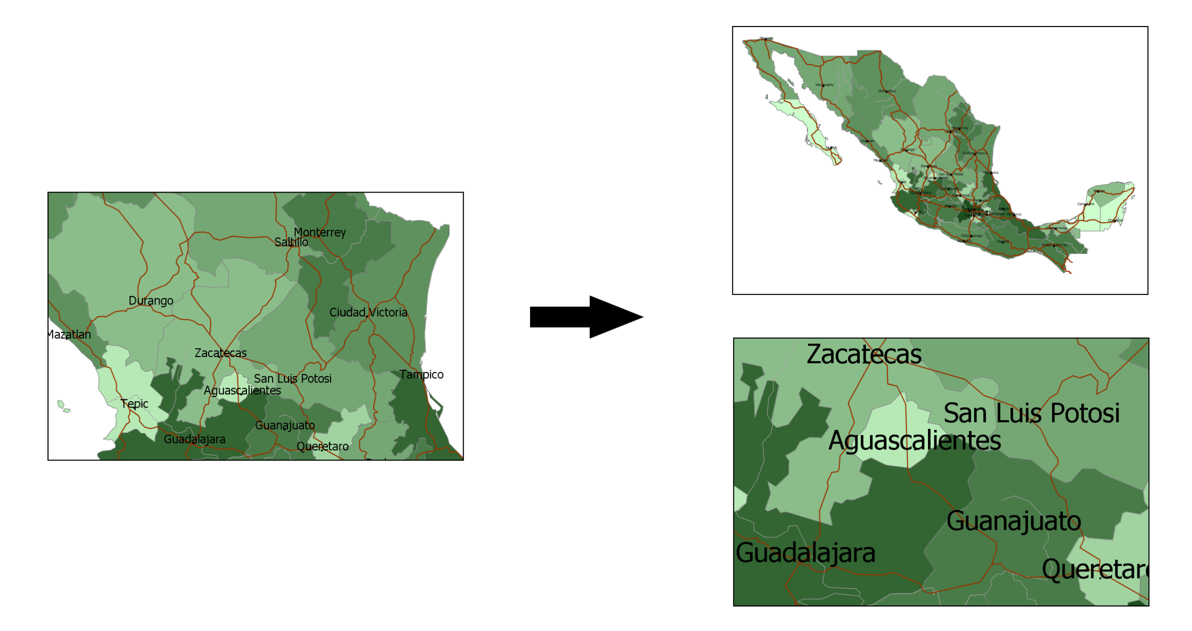


Figura 29.9: El cambio de escala varía el tamaño de los símbolos tales como las etiquetas, haciéndolos demasiado pequeños (a) o demasiado grandes (b)

bles del éxito y la popularidad del SIG 3D es su gran atractivo visual. La tercera dimensión hace más sencillo interpretar buena parte de la información representada, ya que permite mostrarla de un modo más asequible y fácil de entender, especialmente para el observador no especializado. Frente al mapa impreso o la representación bidimensional en pantalla, la representación en tres dimensiones resulta mucho más intuitiva y «real». Al ser más natural y cercano a la realidad que se representa, un mapa tridimensional se percibe menos como un elemento simbólico y más como una realidad.

Por todo ello, porque el factor visual es de gran relevancia en los SIG 3D, una adecuada visualización de la información geográfica tiene mucha importancia para poder aprovechar al máximo todas sus posibilidades. Las siguientes son algunas de las ideas que deben considerarse al trabajar con representaciones tridimensionales, junto, por supuesto, todas las que ya hemos detallado para las representaciones 2D habituales:

- **Existencia de distintas formas de perspectiva**. Existen distintas formas de perspectiva para lograr trasladar la realidad tridimensional a la superficie plana del papel o la pantalla. Estas alteran la percepción de las distintos elementos de la imagen, y en algunas aplicaciones es posible escoger la que se desea, con lo cual aparece un nuevo parámetro que modifica la representación y debe ser ajustado convenientemente.
- **Importancia de la posición del observador y los ángulos de visión**. En un mapa plano no existe como tal el concepto de posición del observador. Aumentando o disminuyendo la escala, el efecto producido es similar a alejarse o acercarse al mapa, y al desplazar este y cambiar el encuadre, podemos considerar que el observador se desplaza, pero estos movimientos no afectan a cómo percibimos la información del mapa. Desde la vista cenital que representa un mapa,



Figura 29.10: Representación saturada al representar elementos con tamaño fijo a una escala baja.

apreciamos sin dificultad las dos dimensiones que este contiene, y ello nos permite interpretar el significado de sus distintos elementos. En el caso tridimensional, la posición del observador no viene únicamente definida por una posición y un alejamiento (que resultan en un encuadre y una escala dadas), sino por una serie de ángulos que al modificarse alteran la visión de las variables representadas.

Por ejemplo, para el caso de que existan elementos tridimensionales tales como edificios, una vista de tipo cenital no dejará percibir adecuadamente la elevación de estos. Por el contrario, una capa ráster de temperaturas representada dentro de esa vista tridimensional sobre el terreno se apreciará mejor si nos situamos por encima de ella, de forma que la línea de visión sea perpendicular.

En otros casos, para una única variable es necesario elegir la visualización en función de aquello que queramos mostrar de forma más clara. Si consideramos una capa de líneas (tridimensional, es decir, formada por un conjunto de puntos definidos mediante 3 coordenadas cada uno) que representa la trayectoria de un avión, la vista cenital nos permitirá ver el recorrido de este, pero será difícil apreciar si ha ascendido o descendido durante el vuelo. Una vista de perfil soluciona esto, pero hace complicado apreciar el desplazamiento en el eje perpendicular a la línea de visión, por lo que el recorrido no se conoce con la misma exactitud. Incluso si este puede apreciarse de algún modo (por ejemplo, variando el grosor de la línea cuando el avión se acerca o aleja del observador para así representar la distancia en profundidad), una capa base con un mapa topográfico no se visualizaría apenas desde esa vista de perfil, haciendo imposible saber cuándo en ese recorrido se ha pasado de un país a otro. A diferencia de lo que sucede con un mapa bidimensional, en una vista tridimensional no se aprecian de igual modo todas las dimensiones implicadas en la representación, ya que el soporte (la pantalla) solo posee dos dimensiones.

- **Orden de capas con un significado distinto**. El orden de representación de capas, según vimos en un punto anterior de este capítulo, define la forma en que estas se pintan y la jerarquía

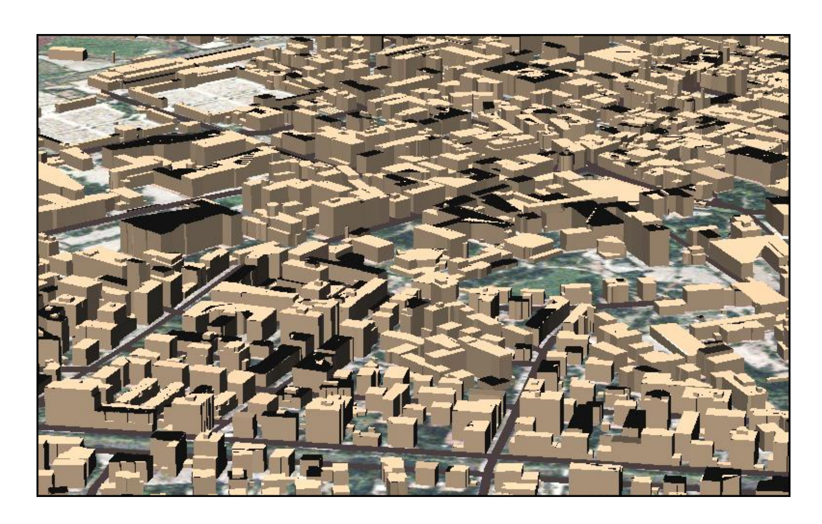


Figura 29.11: La extrusión permite la creación de objetos volumétricos a partir de objetos planos. Los edificios de la imagen se han creado únicamente a partir de la planta y un valor de altura para cada uno de ellos.

que condiciona si la representación de unas capas tapa a la de otras. Se puede considerar como que unas capas se encuentran «encima» de otras. En el caso de una vista 3D, este concepto de «encima» tiene sentido solo si las capas no tienen de por sí una información sobre su altura y se pintan a una altura arbitraria, tal como por ejemplo, sobre el terreno. En caso contrario, será la propia información de la componente z la que dicte dónde se pinta cada capa, y la posición del observador la que condicione cómo se visualizan. En realidad, y salvo para el caso de imágenes que se van a representar a una misma altura y se sobreponen (ya que dentro de la vista 3D ocupan el mismo lugar en el espacio), el concepto de *orden de las capas* no existe como tal cuando trabajamos con una vista 3D.

- **Diferentes formar de incorporar objetos volumétricos**. Existen diversas formas de incluir objetos 3D en una vista tridimensional, la más directa de las cuales es a través de un modelo que defina el objeto a representar. Estos objetos son el elemento adicional que, en el SIG 3D, acompaña a los puntos, líneas y polígonos que conforman las geometrías empleadas en el SIG 2D. Asimismo, se pueden crear geometrías 3D a partir de geometrías 2D como polígonos, mediante el uso de algún atributo asociado a estas y el proceso conocido como extrusión. Dada una capa con la planta de unos edificios (expresada esta con un polígono), y si se conoce la altura de cada uno de ellos, pueden formarse volúmenes (Figura 29.11). Está técnica se emplea habitualmente para la creación de modelos de ciudades cuando no se dispone de modelos detallados de cada edificio. No obstante, cuenta con muchas limitaciones, ya que no permite recrear formas más complejas y no se dispone de información adicional sobre elementos sobre la componente vertical, sino tan solo de la planta, por lo que el alzado carece de detalle (es decir, esos edificios así recreados no tendrán, por ejemplo, ni puertas ni ventanas).
- **La dimensión vertical puede considerarse como otra variable visual alternativa**. En relación con lo comentado en el punto anterior, pueden crearse objetos volumétricos mediante extrusión sin que la dimensión vertical de estos



Figura 29.12: La dimensión vertical puede emplearse como variable visual para visualizar la componente temática de la información geográfica.

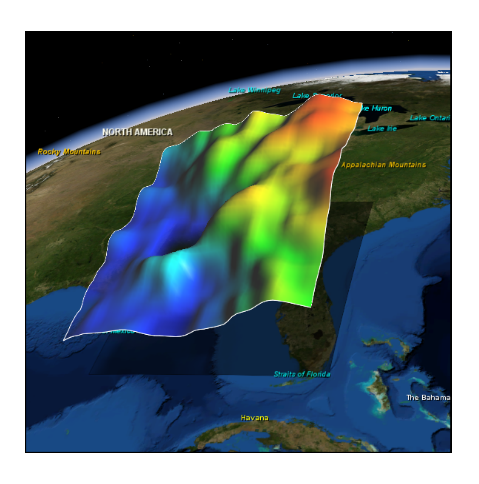


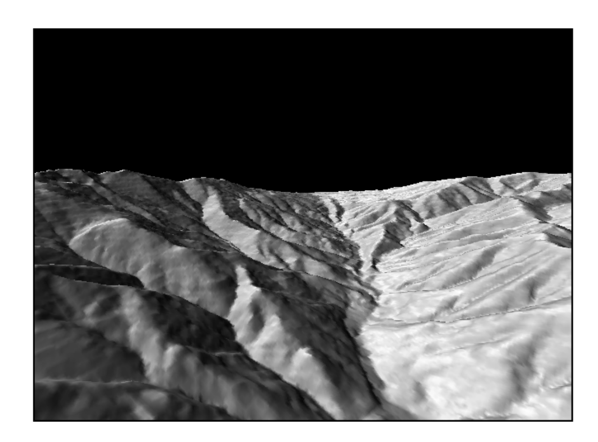
Figura 29.13: La dimensión vertical puede utilizarse también para simbolizar capas ráster con variables distintas a la elevación.

represente necesariamente una altura como tal, sino que esté en función de un parámetro adicional. La figura 29.12 muestra un ejemplo de esto. En la capa visualizada en la imagen, que representa la población de una serie de estados, se ha empleado la elevación para visualizar esta variable, además de recurrir a la habitual gama de valores de colores. Se trata de un mapa de coropletas en el que, sin embargo, no se acude únicamente a la variable visual color para simbolizar la componente temática.

En realidad, estamos utilizando esta junto a la variable tamaño, una variable que para el caso de polígonos no existía en la representación bidimensional (existe, pero debe distorsionarse el contorno del polígono, algo que no resulta adecuado ya que este tiene un significado geográfico, o bien puede aplicarse sobre el grosor de la línea de contorno, lo cual no es tan fácil de percibir e interpretar).

Un planteamiento similar se puede aplicar a capas ráster, como se observa en la figura 29.13. La superficie mostrada sobre el terreno no es un relieve procedente de una capa de elevación, sino de una variable distinta (por ejemplo, presión o temperatura del aire), la cual, además de simbolizarse mediante una rampa de colores, se representa en forma de relieve para hacer más evidente la variación de esos valores. La capa no tiene componente vertical, ya que es una capa ráster bidimensional, por lo que podemos utilizar esa tercera dimensión como variable visual. Hemos visto algunas visualizaciones así en otras partes del libro, por ejemplo en la figura 20.6.

- **Exageración del relieve.** Es habitual que en una visualización tridimensional exista algún modo de distorsionar el relieve para hacerlo más acusado. Mientras que las componentes x e y son proporcionales, la componente z puede alterarse aplicándole un factor de escala para lograr que resulte más obvia la configuración del relieve (Figura 29.14). Esto sirve para acentuar la morfología del terreno, pero también puede ayudar a la comprensión de otras variables, especialmente si el relieve tiene influencia en ellas. Esta exageración se aplica al propio relieve terrestre (es decir, al relieve de un terreno real), así como al que puedan tener las distintas capas debido a la forma en que se representan, tal como en el ejemplo presentado en el punto anterior.



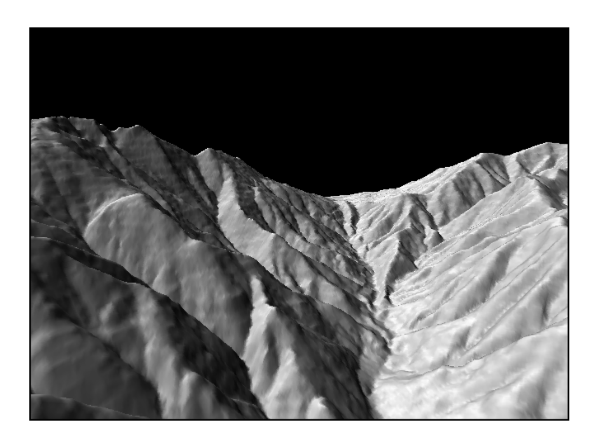


Figura 29.14: La exageración del relieve permite hacer más evidente la configuración de este.

Puede verse en lo anterior la necesidad de extender las ideas del diseño cartográfico para considerar las peculiaridades de las vistas 3D, ya que si no se tienen estas en cuenta, los conceptos de la cartografía clásica, aunque imprescindibles igualmente en este caso, resultan insuficientes. Más información sobre principios de diseño cartográfico para vistas 3D puede encontrarse en [334].

29.7. Visualización dinámica

Mientras que un mapa impreso contiene una información estática que no varía y que representa el estado de unas determinadas variables en un instante dado, dentro de un SIG podemos crear representaciones que vayan variando para mostrarnos la evolución de esas variables. En un SIG es posible no solo visualizar una realidad, sino también el cambio que se produce en esa realidad. Esta visualización dinámica supone una herramienta de gran valor, especialmente para explorar la relación entre distintas variables y cómo el cambio de una de ellas afecta a las restantes.

La visualización dinámica se obtiene mediante una *animación*, la cual se compone de una serie de *escenas*, del mismo modo que una película se compone de una serie de fotogramas. El mapa clásico representa una única de esas escenas, por lo que las nuevas posibilidades que una animación aporta con respecto a este son notables. Aunque de manera distinta a la de una vista tridimensional, una animación aporta también al mapa una dimensión adicional.

El cambio que una animación muestra no ha de darse necesariamente a lo largo del tiempo, sino que puede ser en el espacio o a medida que varía cualquier otra variable. Por ejemplo, una animación puede consistir en un trayecto a lo largo del cual se desplaza el observador y mostrar un «vuelo» entre dos puntos y cómo varía la realidad representada a medida que nos movemos. Este tipo de animaciones son muy comunes en los visores tridimensionales, que permiten definir el trayecto y los parámetros que establecen cómo en los distintos puntos de este el observador mira al terreno.

Podemos, asimismo, escoger cualquier variable adicional como eje de la animación, no es necesario utilizar para ello el tiempo o el espacio.

Al la hora de crear una animación, debemos tener en cuenta no solo las seis variables visuales que estudiamos en el capítulo 27, sino otras seis nuevas, las denominadas *variables visuales dinámicas*[335]:

- **Momento**. El equivalente a la variable visual posición, indica el momento en la animación en que se produce un cambio de una escena a otra.
- **Frecuencia**. Indica la velocidad a la que se produce el cambio en la animación. Si es demasiado lenta, puede aportar una longitud excesiva a esta, mientras que si es demasiado rápida puede hacer difícil analizar e interpretar el cambio que se produce.
- **Duración**. El tiempo que cada escena se encuentra visible, que no tiene que ser el mismo para todas ellas.
- **Magnitud del cambio**. Indica cuánto cambia una escena respecto a la anterior. Si es pequeño, la animación sera fluida, mientras que si es muy elevado, la animación tendrá saltos bruscos. Dividido por la duración nos indica la *tasa de cambio*.
- **Orden**. La posición de cada escena dentro del conjunto, estableciendo antes o después de cuáles de las restantes aparece.
- **Sincronización**. Si la animación muestra la variación de varias variables, establece cómo el cambio en estas se encuentra relacionado. Una correcta sincronización ayuda a interpretar la relación que puede existir entre las variables que varían en la animación.

En un entorno de visualización dinámica, el usuario pueden interactuar también con la representación dinámica, alterando las características de la animación del mismo modo que en una representación estática dentro de un SIG puede modificar el encuadre haciendo uso de las herramientas de navegación habituales.

29.8. Otros elementos de visualización

Además de permitir una representación distinta de los elementos clásicos del mapa y de las variables habituales, la visualización en un SIG puede ampliarse incorporando otros tipos de información distintos, que no tienen presencia en la cartografía tradicional. El ordenador es un soporte más potente que el mapa y es capaz de trabajar además con otros elementos no visuales, de tal modo que ofrece múltiples formas de enriquecer cualquier representación.

En este sentido, el mapa puede comportarse no ya como un documento que trasmite un tipo particular de información (la de tipo geográfico), sino como un contenedor de muchas clases diferentes de información, todas ellas compartiendo el hecho de que pueden localizarse y posicionarse. El mapa se convierte en el elemento de referencia desde el que acceder a todas ellas. Esta es una de las consecuencias del papel que los SIG han jugado en hacernos ver la importancia de la información espacial asociada a la mayoría de fenómenos, hasta el punto de que esa información geográfica, al ser puesta en un mapa, no constituye el objeto primordial de atención, sino que es una información común a otros muchos tipos de información y actúa como nexo de estos.

Algunos de los nuevos elementos que pueden añadirse a una representación en un SIG son los siguientes:

- **Fotografías**. Aunque un mapa puede contener fotografías, está muy limitado en temas de espacio, y la presencia de estas es anecdótica. Por el contrario, y gracias a sus elementos interactivos, un SIG puede incorporar fotografías solo a una determinada escala, y solo si el usuario así lo pide, por ejemplo haciendo clic en un símbolo concreto. Esto permite incorporar un número ilimitado de imágenes, permitiendo que



Figura 29.15: Las fotografías inmersivas permiten al usuario «meterse» en el mapa, ampliando la información que se muestra acerca de un lugar con imágenes reales tomadas sobre el terreno. Al igual que se navega por un mapa, el usuario puede navegar por el terreno haciendo uso de los controles interactivos correspondientes (imagen tomada de Google Street View).

estas complementen a la información visual del propio mapa.

Un caso particular son las fotografías de tipo inmersivo, en las que el usuario puede navegar a través de fotografías del entorno como si se encontrara realmente en él (Figura 29.15).

- **Vídeos**. Del mismo modo que las fotografías, aportan más información sobre la zona representada y permiten una exploración más profunda. Aunque son una tecnología aún muy experimental, existen también vídeos de tipo inmersivo.
- **Sonido**. Los elementos no han de ser necesariamente visuales, sino que pueden proporcionar información a través de otros sentidos distintos. Es posible, por ejemplo, añadir elementos sonoros a un mapa para facilitar su uso a personas con baja visión. Utilizando síntesis de voz, la aplicación SIG puede leer en voz alta las etiquetas de un mapa o los valores asociados a los distintos elementos de este.
- **Documentos**. Un SIG puede incorporar documentos complejos tales como páginas Web o textos varios.

Esta lista, no obstante, es muy susceptible de extenderse, ya que, virtualmente, un SIG puede incorporar cualquier elemento que pueda manejarse dentro de un ordenador. Cada día aparecen nuevas ideas sobre cómo combinar la información geográfica con otros tipos de información, y el SIG se sitúa en la base de todos estos nuevos planteamientos como herramienta fundamental de trabajo.

29.9. Resumen

Hemos visto en este capítulo cómo aplicar las ideas de capítulos previos a la representación de capas en un SIG. Cada tipo de capa tiene sus particularidades, y es en función de estas que se debe buscar la mejor forma de emplear las variables

visuales y los conceptos de simbología gráfica que ya conocemos.

Una parte de las representaciones que generamos en un SIG están destinadas a ser representadas en pantalla, y es necesario considerar este hecho a la hora de visualizar información geográfica en un SIG. Dos son los principales aspectos que han de tenerse en cuenta: la baja resolución de la pantalla en comparación con el papel y la interactividad propia de la representación.

Además de trabajar con las formas cartográficas clásicas que un SIG es capaz de producir, existen nuevas formas, entre las que destacan las vistas tridimensionales y las animaciones. Junto a ellas, una de las tendencias actuales que aumentan las posibilidades de un SIG como herramienta de visualización es la incorporación de otros elementos tales como vídeos, fotografías u otros documentos de diversas clases.

Bibliografía

[1] C.D. Tomlin. *Geographic information systems and cartographic modelling.* Prentice Hall., 1990.

[2] J. Star and J. Estes. *Geographic Information Systems: An Introduction.* Prentice-Hall, 1990.

[3] G. Korte. *The GIS Book (5th Ed. Rev.).* Autodesk Press, 2001.

[4] M. F. Goodchild. What is geographic information science? ncgia core curriculum in giscience, 1997.

[5] ESRI. Gis and cad-the right tool for the job. an esri white paper. Technical report, ESRI, 2002.

[6] J.A. Báguena. El geomarketing y los sistemas de información para la toma de decisiones. *Mapping interactivo*, 26, 1995.

[7] ESRI. Enterprise geographic information servers: A new information system architecture. a esri white paper. Technical report, ESRI, 2003.

[8] M. Goodchild. Citizens as sensors: the world of volunteered geography, 2007.

[9] T.W. Foresman. *The history of Geographic Information Systems.* Prentice Hall, 1998.

[10] F. A Sarria. Sistemas de informacion geografica. 2005.

[11] I. Heywood. *An introduction to Geographical Information Systems.* Longman, 1998.

[12] W. Clinton. Executive order 12906: Coordinating geographic data acquisition and access: The national spatial data infrastructure. *Federal Register*, 59:17671–17674, 1994.

[13] M. Craglia. Inspire:the european spatial data infrastructure. In *INSPIRE Conference 2009*, 2009.

[14] D. Martin. *Geographical Information Systems and their Socio-economic Applications.* Routledge, 1991.

[15] I.S. Evans. General geomorphometry, derivatives of altitude, and descriptive statistics. In R.J. Chorley, editor, *Spatial analysis in geomorphology*, pages 17–90. Harper & Row, 1972.

[16] P. Whittle. On stationary processes in the plane. *Biometrika*, 41:434–49, 1954.

[17] M. Goodchild and R. Haining. Gis and spatial data analysis: Converging perspectives. *Journal of Economics*, 83(1):363–385, October 2003.

[18] F.P. Preparata and M.I. Shamos. *Computational Geometry - An Introduction.* Springer-Verlag, 1985.

[19] F. Martín Ass. *Geodesia y Cartografía Matemática.* IGN, 1983.

[20] G. D. Jenerette and J. Wu. On the definitions of scale. *ESA Bulletin*, 81:104–105, 2000.

[21] N. S-N. Lam and D. A. Quattrochi. On the issues of scale, resolution, and fractal analysis in the mapping sciences. *Professional Geographer*, 44, 1992.

[22] Anon. *Generalización cartográfica.* EUITP, 2002.

[23] R.B. McMaster and K.S. Shea. *Generalization in digital cartography.* Association of American Geographers, 1992.

[24] R. Diener. Information science: What is it?...what should it be? *Bulletin of ASIS*, June/July:17–21, 1989.

[25] Information science: definition and scope. In J. Williams and Carbo T., editors, *Information Science: Still an Emerging Discipline.* Cathedral Publishing, 1997.

[26] D.F. Sinton. The inherent structure of information as a constraint to analysis: Mapped thematic data as a case study. In *Harvard Papers on Geographic Information Systems, vol. 6*, pages 1–17. Addison-Wesley, 1978.

[27] H. Couclelis. People manipulate objects (but cultivate fields): beyond the raster-vector debate in gis. In Frank A. U., I. Campari, and U. Formentini, editors, *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space, vol. 639, Lecture Notes in Computer Science*, pages 65–77. Springer-Verlag, 1992.

[28] M.F. Goodchild and K.K. Kemp. Ncgia core curriculum in gis, 1990.

[29] B. Diaz and S. Bell. *Spatial Data Processing using Tesseral Methods.* Natural Environment Research Council, Reading, UK, 1986.

[30] J.R. Herring. Tigris: Topologically integrated geographical information systems. In *Proceedings, AutoCarto 8*, 1987.

[31] J. Husdal. How to make a straight line square - network analysis in raster gis. Master's thesis, University of Leicester, UK, 2000.

[32] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin. *Network Flows: Theory, Algorithms and Applications.* 1993.

[33] T.K Peucker, R.J. Fowler, J.J. Little, and D.M. Mark. The triangulated irregular network. In *Proceedings, American Society of Photogrammetry: Digital Terrain Models (DTM) Symposium, St. Louis, Missouri, May 9-11,*, pages 516–540, 1978.

[34] D. M. Mark. Computer analysis of topography: A comparison of terrain storage methods. *Geografisker Annaler*, 57A:179–188, 1975.

[35] M. Egenhofer and J. Herring. *Geographical Information Systems, Vol. 1: Principles*, chapter High-Level Spatial Data Structures, pages 227–237. Longman, 1991.

[36] G.M. Morton. A computer oriented geodetic data base; and a new technique in file sequencing. Technical report, IBM Ltd, 1966.

[37] R. Finkel and J.L. Bentley. Quad trees: A data structure for retrieval on composite keys. *Acta Informatica*, 4(1):1–9, 1974.

[38] R. H. Guting. An introduction to spatial database systems. *VLDB Journal*, 3(4), 1994.

[39] A. Buchmann, O. Gunther, T.R. Smith, and Y.-F. Wang. *Design and Implementation of Large Spatial Databases.* Number 409 in Unit Notes in Computer Science. 1990.

[40] M.F. Worboys and M. Duckham. *GIS: A Computing Perspective, Second Edition.* CRC Press, 2004.

[41] M.J. Jackson and P.A. Woodsford. *Geographical Information Systems: Principles*

and Applications, chapter GIS data capture hardware and software, pages 239–249. Longman, 1991.

[42] P. Curran. *Principles of remote sensing*. Longman, 1989.

[43] M.Tt. Lillesand and W. R. Kiefer. *Remote Sensing and Image Interpretation, 3rd Edition.* John Wiley & Sons, Inc., 1997.

[44] E. Chuvieco. *Fundamentos de teledetección espacial 3 ed.* Ediciones Rialp, 1996.

[45] S. Reutebuch Andersen, H-E and G. Schreuder. Automated individual tree measurement through morphological analysis of a lidar-based canopy surface model. In *Proceedings of the First International Precision Forestry Symposium*, pages 11–22, 2001.

[46] K.Kraus and 2001 N. Pfeifer. Advanced dtm generation from lidar data. In *IASPRS*, volume XXXIV. 2001.

[47] r. Welch and T.R. Jordan. *Raster Imagery in Geographic Information Systems*, chapter Using scanned air photographs, pages 55–69. Onward Press, 1996.

[48] C. A. Davis, F. T. Fonseca, and K. A. V. Borges. A flexible addressing system for approximate geocoding. In *V Brazilian Symposium on GeoInformatics*, 2003.

[49] H. Bonneval. *Photogrammetrie Generale*. Editions Eyrolles, 1972.

[50] Dod world geodetic system 1984, its definition and relationship with local geodetic systems. Technical report, National Imagery and Mapping Agency, 1997.

[51] F. Van Graas and M. Braasch. Gps interferometric attitude and heading determination: Initial flight test results. *Navigation*, 38:297–316, 1991.

[52] C.Torrecillas Lozano and J.J. Martínez. Resolucion de ambigüedades gps: técnicas empleadas y estudios futuros. *Mapping Interactivo*, 49, 1998.

[53] B.W Remondi. Kinematic and pseudo-kinematic gps. In *Proceedings of the Satellite Division Technical Meeting. The Institute of Navigation.*, pages 115–121, 1988.

[54] Chris Rizos and Shaowei Han. Precise kinematic applications of gps: prospects and challenges. *B. Ci. Geodésicas*, 3:34–40, 1998.

[55] L. Asenjo. *Fundamentos para la realización de levantamientos mediante el sistema GPS.* E.T.S.C.G.T. Universidad Politécnica de Valencia, 1997.

[56] Karen Steede-Terry. *Integrating GIS and the Global Positioning System.* ESRI Press, 2000.

[57] M.F. Goodchild. Citizens as voluntary sensors: Spatial data infraestructure in the world of web 2.0. *International Journal of Spatial Data Infraestructures Research*, 2:24–32, 2007.

[58] O. Hellwich. Sar interferometry: Principles, processing, and perspectives.

[59] K. G. Nikolakopoulos, E. K. Kamaratakis, and N. Chrysoulakis. Srtm vs aster elevation products. comparison for two regions in crete, greece. *International Journal of Remote Sensing*, 27(21):4819–4838, 2006.

[60] J. Bryan Mercer and S. Schnick. Comparison of dems from star-3i interferometric

sar and scanning laser. In *Proceedings of the ISPRS WG III/2,5 Workshop: Mapping Surface Structure and Topography by Airborne and Spaceborne Lasers,*, 1999.

[61] J. Bryan Mercer. Comparing lidar and ifsar: What can you expect? In *Proceedings, Photogrammetric Week*, 2001.

[62] Une-en-iso 8402 gestión de la calidad y aseguramiento de calidad.vocabulario (iso 8402:1994). Technical report, AENOR, 1995.

[63] M. Gómez Delgado and J. Bosque Sendra. Aplicación de análisis de incertidumbre como método de validación y control del riesgo en la toma de decisiones. *GeoFocus*, 4:179–208, 2004.

[64] Kenneth E. Foote and Donald J. Huebner. Error, accuracy, and precision, 2000.

[65] Pepijn van Oort. *Spatial data quality: from description to application.* Netherlands Geodetic Commission, 2005.

[66] P.F. Fisher. *Geographical Information Systems: Principles and technical issues*, chapter Models of uncertainty in spatial data, pages 191–205. John Wiley & Sons, 1999.

[67] G. B. M. Heuvelink. *Error propagation in environmental modeling with GIS.* Taylor & Francis, 1998.

[68] V. Barnett and T. Lewis. *Outliers in Statistical Data.* Wiley, 1995.

[69] M. Last and A. Kandel. Automated detection of outliers in real-world data. In *Proceedings of the Second International Conference on Intelligent Technologies (InTech 2001)*, pages 292–301, 2001.

[70] Chang-Tien Lu, Dechang Chen, and Yufeng Kou. Algorithms for spatial outlier detection. In *Proceedings of the Third IEEE International Conference on Data Mining*, 2003.

[71] Shashi Shekhar, Chang-Tien Lu, and Pusheng Zhang. A unified approach to detecting spatial outliers. *GeoInformatica*, 7(2), 2003.

[72] T. Hengl and H.I. Reuter, editors. *Geomorphometry: Concepts, Software, Applications*, volume 33. Elsevier, Amsterdam, 2008.

[73] N. Chrisman. Error in categorical maps: Testing versus simulation. In *Proceedings, AutoCarto 9*, pages 521–529, 1989.

[74] J. D. Vitek, S. J. Walsh, and M. S. Gregory. Accuracy in geographic information systems: an assessment of inherent and operational errors. *Pecora*, 9:296–302, 1984.

[75] S. Walsh, D. Lightfoot, and D. Butler. Recognition and assessment of error in gis. *Photogrammetric Engineering and Remote Sensing*, 52:1423–1430, 1987.

[76] J.A. Newcomer and J. Szajgin. Accumulation of thematic map errors in digital overlay analysis. *The American Cartographer*, 11(1):58–62, 1984.

[77] U.S Department of Transportation. *Data Integration Glossary.* 2001.

[78] E.F. Codd. A relational model of data for large shared data banks. *Communications of the ACM*, 13(6):377–387, 1970.

[79] M. Marqués. *Bases de Datos orientadas a objetos.* 2002.

[80] D. Hart and H.Philips. The metadata primer.

[81] L. Anselin. What is special about spatial data? alternative perspectives on spatial data analysis. In *Symposium on Spatial Statistics, Past, Present and Future*. Department of Geography, Syracuse University., 1989.

[82] R.J.f Chorley. *Spatial Analysis in Geomorphology*. Methuen, 1972.

[83] T.L. Nyerges. Analytical map use. *Cartography and GIS*, 18, 1991.

[84] F. Slater. *Learning through Geography*. Heineman Educational Books, Ltd, 1982.

[85] A. Rosenfeld and Thurston M. Edge and curve detection for visual scene analysis. *IEE Trans. on Computers*, 20-c(5):562–569, 1971.

[86] B.B. Mandelbrot. *The Fractal Geometry of Nature*. W.H. Freeman and Company, 1982.

[87] S Openshaw. *The Modifiable Areal Unit Problem*. Concepts and Techniques in Modern Geography 38. GeoBooks., 1983.

[88] I Heywood. *Introduction to Geographical Information Systems*. Addison Wesley Longman., 1998.

[89] T. Nakaya. An information statistical approach to the modifiable areal unit problem in incidence rate maps. *Environment & Planning*, 32, 2000.

[90] H.D. Reynolds. *The modifiable area unit problem: empirical analysis by statistical simulation*. PhD thesis, University of Toronto, 1998.

[91] W.R. Tobler. A computer movie simulating urban growth in the detroit region. *Economic Geography*, 46(2):234–240, 1970.

[92] S. Openshaw and P. Taylor. *Statistical Applications in the Spatial Sciences*, chapter A million or so correlation coefficients: three experiments on the modifiable areal unit problem., pages 127–144. Pion, 1979.

[93] J. Chen. A raster-based method for computing voronoi diagrams of spatial objects using dynamic distance transformation. *Journal of Geographical Information Science*, 13(3):209–225, 2001.

[94] G. Borgefors. Distance transformation in digital images. *Computer vision, graphics and image processing*, 34:344–371, 1986.

[95] E. Haines. *Graphics Gems IV*, chapter Point in Polygon Strategies, pages 24–46. Academic Press, 1994.

[96] D. V. Pullar and M. J. Egenhofer. Toward formal definitions of topological relations among spatial objects. In *Proceedings Third International Symposium on Spatial Data Handling, Sydney, Australia, August 17-19*, pages 225–241, 1988.

[97] D. Hernandez. *Qualitative representation of Spatial Knowledge*. Springer-Verlag, 1994.

[98] Jun Xu. Formalizing natural-language spatial relations between linear objects with topological and metric properties. *International Journal of Geographical Information Science*, 21(4):377–395, 2007.

[99] A.U. Frank and D.M. Mark. *Geographical Information Systems: Principles and Applications*, chapter Language issues for GIS, pages 147–163. Longmans, 1991.

[100] A. R. Boyle, J. Dangermond, D. F. Marble, D. S. Simonett, L. K. Smith, and R. F. Tomlinson. Final report of a conference on the review and synthesis of problems and

directions for large scale geographic information system development. Technical report, National Aeronautics and Space Administration, 1983.

[101] D. M. Mark and M. J. Egenhofer. Modeling spatial relations between lines and regions: Combining formal mathematical models and human subjects testing. *Cartography and Geographic Information Systems*, 21(4):195–212, 1994.

[102] M. A. Egenhofer. Formal definition of binary topological relationships. In *Third international conference on data organization and algorithms (FODO) Lecture notes in computer science*, pages 457–472. Springer-verlag, 1989.

[103] Max J. Egenhofer and A. Rashid B. M. Shariff. Metric details for natural-language spatial relations. *ACM Transactions on Information Systems*, 16(4):295–321, 1998.

[104] K.A. Nedas, M.J. Egenhofer, and D. Wilmsen. Metric details of topological line-line relations. *International Journal of Geographical Information Science*, 21(1):21–48, 2007.

[105] A. Beaulieu. *Aprende SQL*. Anaya, 2003.

[106] M.J. Egenhofer. Spatial sql: a query and presentation language. *IEEE Transactions on Knowledge and Data Engineering*, 6:86–95, 1994.

[107] C. Freksa. *Cognitive and linguistic aspects of geographic space*, chapter Qualitative Spatial Reasoning. Kluwer, 1991.

[108] J. P. Cole and C. A. M. King. *Quantitative Geography*. John Wiley & Sons, 1968.

[109] L. J. King. *Statistical Analysis in Geography*. Prentice-Hall, 1962.

[110] Peter A. Rogerson. *Statistical Methods for Geography*. Sage, 2001.

[111] R. Bachi. Standard distance measures and related methods for spatial analysis. *Papers, Regional Science Association*, 10:83–132., 1963.

[112] C. McGrew, J. & Monroe. *An introduction to statistical problem solving in geography*. William C. Brown publishers, 1993.

[113] E. Batchelet. *Circular statistics in biology*. Academic Press., 1981.

[114] N. I. Fisher. *Statistical analysis of circular data*. Cambridge University Press., 1993.

[115] N.H. Kuiper. Test concerning random points on a circle. *Proc K. Ned Akad Ser A: Math Sci 63*, 63:38–47, 1960.

[116] J.S. Rao. *Some contributions to the analysis of circular data*. PhD thesis, Indian Statistical Institute, Calcutta., 1969.

[117] G.S. Russell and D.J. Levitin. An expanded table of probability values for rao's spacing test. *Communications in statistics: Simulation and computation*, 24:879–888, 1995.

[118] N.A.C. Cressie. *Statistics for spatial data*. John Wiley & Sons, 1991.

[119] F. David and P. Moore. Notes on contagious distributions in plant populations. *Annals of Botany*, 18:47–53, 1954.

[120] J. Douglas. Clustering and aggregation. *Sankhya*, 37:398–417., 1975.

[121] M. Morisita. Measuring of the dispersion and analysis of distribution patterns. *Memoires of the Faculty of Science, Kyushu University*, 2:215–235., 1959.

[122] P. J. Evans and F. C. Evans. Distance to nearest neighbour as a measure of spatial relations in populations. *Ecology*, 35:445–453, 1954.

[123] K. P. Donnelly. Simulations to determine the variance and edge efect of total nearest neighbour distance. In I. Hodder, editor, *Simulation Studies in Archaeology*, pages 91–95. Cambridge University Press, 1978.

[124] B.D. Ripley. Modelling spatial patterns. *Journal of the Royal Statistical Society. Series B 30*, pages 172–192, 1977.

[125] D. R. Cox and V. Isham. *Point Processes*. Chapman & Hall, 1980.

[126] L. Anselin. Spatial data analysis with gis: an introduction to application in the social sciences, technical report 92-10. Technical report, NCGIA, 1992.

[127] P. Moran. *The interpretation of statistical maps*. Journal of the Royal Statistical Society B 10, 24351., 1948.

[128] R. Geary. The contiguity ratio and statistical mapping. *The Incorporated Statistician*, 5:115–145, 1954.

[129] A.D. Cliff and JK Ord. *Spatial Autocorrelation*. Pion, 1973.

[130] A. Getis and J.K. Ord. The analysis of spatial association by use of distance statistics. *Geographical Analysis*, 24:189–206, 1992.

[131] L. Anselin. The local indicators of spatial association: Lisa. *Geographical Analysis*, 27:93–115, 1995.

[132] E.H. Isaaks and R.M. Srivastava. *Applied Geostatistics*. Oxford University Press, New York, 1989.

[133] P. Goovaerts. *Geostatistics for Natural Resources Evaluation (Applied Geostatistics)*. Oxford University Press, New York, 1997.

[134] P.A. Burrough. *Principles of Geographic Information Systems for Land Resource Assessment*. Monographs on Soil and Resources Survey No. 12. Oxford Science Publications, New York, 1986.

[135] Donald Shepard. Two-dimensional interpolation function for irregularly-spaced data. In *Proc ACM National Conference*, pages 517–524., 1968.

[136] D. G Schweikert. An interpolation curve using a spline in tension. *Journal of Mathematics and Physics,*, 45:312–313, 1966.

[137] H. Mitasova and J. Hofierka. Interpolation by regularized spline with tension, ii application to terrain modelling and surface geometry analysis. *Mathematical Geology*, 25:657–669, 1993.

[138] G. Matheron. Principles of geostatistics. *Economic Geology*, 58:1246–1266, 1963.

[139] D.G. Krige. A statistical approach to some mine valuations and allied problems at the witwatersrand. Master's thesis, University of Witwatersrand, 1951.

[140] R.H. Myers. *Classical and Modern Regression with Applications*. PWS-Kent Publishing, 1990.

[141] R. Haining. *Spatial Data Analysis: theory and practice*. Cambridge University Press, 2003.

[142] R. Sibson. *Interpolating multivariate data*, chapter A Brief Description of Natural Neighbor Interpolation, pages 21–36. John Wiley & Sons, 1981.

[143] Waldo R. Tobler. Smooth pycnophylactic interpolation for geographical regions. *Journal of the American Statistical Association*, 74(367):519–530, 1979.

[144] T. Hengl. Finding the right pixel size. *Computers & Geosciences*, 32(9):1283–1298, 2006.

[145] B.W. Avery. Soil survey methods: a review. technical monograph no. 18. Technical report, Soil Survey & Land Resource Centre, Silsoe., 1987.

[146] B.D. Silverman. *Density estimation for statistics and data analysis.* Chapman and Hall, 1986.

[147] W.H. Wischmeier and D.D. Smith. *Predicting Rainfall Erosion Losses. Agriculture Handbook 537.* United States Department of Agriculture. Science and Education Administration., 1978.

[148] R. R Yager. On ordered weighted averaging aggregation operators in multi-criteria decision making. *IEEE Trans. on SMC*, 18:183–190, 1988.

[149] USDA. Technical release 55 (tr-55) urban hydrology for small watersheds, second edition. Technical report, Natural Resources Conservation Service, Conservation Engineering Division., 1986.

[150] D R. Caldwell. Extending map algebra with flag operators. In *Proceedings GeoComputation 2000*, 2000.

[151] J. Wood. *The Geomorphological Characterisation of Digital Elevation Models.* Ph.d.thesis, University of Leicester, 1996.

[152] K. McGarigal, S.A. Cushman, Neel M.C., and E. Ene. Fragstats: Spatial pattern analysis program for categorical maps. computer software program produced by the authors at the university of massachusetts, amherst, 2002.

[153] K. Turkowski. Filters for common resampling tasks. In *Graphic Gems*, pages 147–165. Academic Press Professional,Inc, 1990.

[154] G. Camara, D. Palomo, and R.C. Souza. Towards a generalized map algebra: principles and data types. In *Proceeding VII Simposio Brasileiro de GeoInformatica.*, 2005.

[155] M. Takeyama and H. Couclelis. Map dynamics: integrating cellular automata and gis through geo-algebra. *International Journal of Geographical Information Science*, 11:73–91, 1997.

[156] G.X. Ritter, J. Wilson, and J. Davidson. Image algebra: An overview. *Computer Vision, Graphics and Image Processing*, 49:297–331, 1990.

[157] R.J. Pike. Geomorphometry — progress, practice, and prospect. *Zeitschrift für Geomorphologie*, 101:221–238, 1995.

[158] R.J. Pike. Geomorphometry — diversity in quantitative surface analysis. *Progress in Physical Geography*, 24(1):1–20, 2000.

[159] S. Rasemann, J. Schmidt, L. Schrott, and R. Dikau. Geomorphometry in mountain terrain. In M. P. Bishop and J. F. Shroder, editors, *GIS & Mountain Geomorphology*, pages 101–145. Springer, Berlin, 2004.

[160] A.M. Felicísimo. *Modelos digitales del terreno. Introducción y aplicaciones en las ciencias ambientales.* Pentalfa Ediciones, 1994.

[161] Severn tidal tributaries catchment flood management plan scoping stage. Technical report, UK: Environment Agency., 2005.

[162] Y. Gyasi-Agyei, G. Willgoose, and F. P. De Troch. Effects of vertical resolution and map scale of digital elevation models on geomorphological parameters used in hydrology. *Hydrol. Processes*, 9:363–382, 1995.

[163] J.P. Walker and G.R. Willgoose. On the effect of dem accuracy on hydrology and geomorphology models. *Wat. Resour. Res*, 35(7):2259–2268, 1998.

[164] A. Sindayihebura, M. Van Meirvenne, and S. Nsabimana. Comparison of methods for deriving a digital elevation model from contours and modelling of the associated uncertainty. In M. Caetano and M. Painho., editors, *Proceedings 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, 2006.

[165] M.F. Hutchinson. A new procedure for gridding elevation and stream line data with automatic removal of spurious pits. *Journal of Hydrology*, 106:211–232, 1989.

[166] A. Felicísimo. Parametric statistical method for error detection in digital elevation models. *ISPRS Journal of Photogrammetry and Remote Sensing*, 49(4):29–33, 1994.

[167] K.H. Jones. A comparison of eight algorithms used to compute slopes as a local property of the dem. In *Proceedings of the GIS Research UK 1996 Conference*, pages 7–12, 1996.

[168] J. Schmidt, I. Evans, and J. Brinkmann. Comparison of polynomial models for land surface curvature calculation. *International Journal Geographical Information Science*, 17:797–814, 2003.

[169] L.W. Zevenbergen and C.R. Thorne. Quantitative analysis of land surface topography. *Earth Surface Processes Landforms*, 12:47–56, 1987.

[170] P.A. Shary. Land surface in gravity points classification by a complete system of curvatures. *Mathematical Geology*, 27(3):373–390, 1995.

[171] B.K.P. Horn. Hill shading and the reflectance map. *Proceedings IEEE*, 69(1):14–47, 1981.

[172] D.B. Beasley and L.F Huggins. *ANSWERS: User's manual. U.S. EPA-905/9-82-001.* U.S. EPA, Chicago, IL, 1982.

[173] M.R. Travis, G.H. Elsner, W.D. Iverson, and C.G. Johnson. *VIEWIT: computation of seen areas, slope, and aspect for land-use planning.* USDA, Berkeley, California, 1975.

[174] J. K. Berry. *Beyond Mapping: Concepts, Algorithms, and Issues in GIS.* John Wiley & Sons, 1996.

[175] R.D. Hobson. Surface roughness in topography: a quantitative approach. In R.J. Chorley, editor, *Spatial analysis in geomorphology*, pages 221–245. Harper & Row, 1972.

[176] J.S. Jenness. Calculating landscape surface area from digital elevation models. *Wildlife Society Bulletin*, 32(3):829–839, 2004.

[177] J.C. Gallant and J.P. Wilson. Tapes-g: a grid-based terrain analysis program for the environmental sciences. *Computers & Geosciences*, 22(7):713–722, 1996.

[178] P.A. Shary, L.S. Sharaya, and A.V. Mitusov. Fundamental quantitative methods of land surface analysis. *Geoderma*, 107(1-2):1–32, 2002.

[179] D. King, H. Bourennane, M. Isambert, and J.J. Macaire. Relationship of the presence of a non-calcareous clay–loam horizon to DEM attributes in a gently sloping area. *Geoderma*, 89(1-2):95–111, 1999.

[180] R. Koethe and F. Lehmeier. Sara, system zur automatischen relief-analyse,benutzerhandbuch. 1996.

[181] P.A. Shary, L.S. Sharaya, and A.V. Mitusov. The problem of scale-specific and scale-free approaches in geomorphometry. *Geografia Fisica e Dinamica Quaternaria*, 28(1):81–101, 2005.

[182] D. Riaño, E. Chuvieco, J. Salas, and I. Aguado. Assessment of different topographic corrections in landsat-tm data for mapping vegetation types. *IEEE Transactions on Geoscience and Remote Sensing*, 41(5):1056–1061, 2003.

[183] John P. Wilson and John C. Gallant. *Terrain Analysis: Principles and Applications.* John Wiley & Sons, Inc., 2000.

[184] J. Krcho. Morphometric analysis of relief on the basis of geometric aspect of field theory. *Acta Geographica Universitatis Comenianae, Geographico-Physica*, 1(1):7–233, 1973.

[185] J. Krcho. Teoretická concepcia a interdisciplinarne aplikacie komplexného digitalneho modelu reliéfu pri modelovanï dvojdimenzionalnych poli. *Geografický casopis*, 35(3):265–291, 1983.

[186] J. O'Callaghan and D. Mark. The extraction of drainage networks from digital elevation data. *Computer Vision, Graphics, and Image Processing*, 28:323–344, 1984.

[187] J. Fairfield and P. Leymarie. Drainage networks from grid digital elevation models. *Water Resources Research*, 27(5):709–717, 1991.

[188] P. Quinn, K. Beven, P. Chevallier, and O. Planchon. The prediction of hillslope paths for distributed hydrological modeling using digital terrain models. *Hydrological processes*, 5:59–79, 1991.

[189] P. Pilesjo and Q. Zhou. Theoretical estimation of flow accumulation from a grid-based digital elevation model. In *Proceeding Geoinformatics 97 Conference*, 1997.

[190] P Holmgren. Multiple flow direction algorithms for runoff modelling in grid based elevation models: An empirical evaluation. *Hydrological processes*, 8:327–334, 1994.

[191] D.G. Tarboton. A new method for the determination of flow directions and upslope areas in grid digital elevation models. *Water Resources Research*, 33(2):309–319, 1997.

[192] N.L. Lea. An aspect driven kinematic routing algorithm. In A.J. Parsons and A.D. Abrahams, editors, *Overland flow: Hydraulics and erosion mechanics*, pages 147–175. Chapman & Hall, New York, 1992.

[193] M. Costa-Cabral and S.J. Burges. Digital elevation model networks (demon): a model of flow over hillslopes for computation of contributing and dispersal areas. *Water Resources Research*, 30(6):1681–1692, 1994.

[194] P. Pilesjo. Estimation of flow distribution for hydrological modelling. In *Proceeding Geoinformatics 98 Conference*, 1998.

[195] S.K. Jenson and J.O. Domingue. Extracting topographic structure from digital elevation data for geographical information system analysis. *Photogrammetric Engineering and Remote Sensing*, 54(11):1593–1600, 1988.

[196] J. Garbrecht and L. W. Martz. The assignment of drainage direction over flat surfaces in raster digital elevation models. *Journal of Hydrology*, 193(1-4):204–213, 1997.

[197] D. Mark. Automated detection of drainage networks from digital elevation models. *Cartographica*, 21:168–178, 1984.

[198] L. Martz and J. Garbrecht. An outlet breaching algorithm for the treatment of closed depressions in a raster DEM. *Computers and Geosciences*, 25(7):835–844, 1999.

[199] O. Planchon and F. Darboux. A fast, simple and versatile algorithm to fill the depressions of digital elevation models. *Catena*, 46:159–176, 2001.

[200] O. Conrad. *SAGA - Entwurf, Funktionsumfang und Anwendung eines Systems für Automatisierte Geowissenschaftliche Analysen.* PhD thesis, University of Goettingen, 2007.

[201] R.E. Horton. Drainage basin characteristics. *Transactions American Geophysical Union*, 14:350–361, 1932.

[202] A.N. Strahler. Quantitative geomorphology of drainage basins and channel networks. In V.T. Chow, editor, *Handbook of Applied Hydrology*, pages 4–39–4–76. McGraw-Hill, New York, 1964.

[203] D.G. Tarboton, R.L. Bras, and I. Rodriguez-Iturbe. On the extraction of channel networks from digital elevation data. *Hydrologic Processes*, 5(1):81–100, 1991.

[204] A.J. Broscoe. Quantitative analysis of longitudinal stream profiles of small watersheds. Technical report, Office of Naval Research, Project NR 389-042, Technical Report No. 18, Department of Geology, Columbia University, 1959.

[205] D. R. Montgomery and W. E. Dietrich. Source areas, drainage density, and channel initiation. *Water Resources Research*, 25(8):1907–1918, 1989.

[206] S.D. Peckham. Efficient extraction of river networks and hydrologic measurements from digital elevation data. In Barndorff-Nielsen et al., editor, *Stochastic Methods in Hydrology: Rain, Landforms and Floods*, pages 173–203. World Scientific, Singapore, 1998.

[207] D. P. Tarboton, D.G.; Ames. Advances in the mapping of flow networks from digital elevation data. In *World Water and Environmental Resources Congress*, 2001.

[208] T. K. Peuker and D. H. Douglas. Detection of surface specific points by local parallel processing of discrete terrain elevation data. *Computer Graphics and Image Processing*, 4:375–487, 1975.

[209] R. Turcotte, J.P. Fortin, A.N. Rousseau, S. Massicotte, and J.P. Villeneuve. Deter-

mination of the drainage structure of a watershed using a digital elevation model and a digital river and lake network. *Journal of Hydrology*, 240:225–242, 2001.

[210] S. Shreve, R.L. Statistical law of stream numbers. *J. Geol*, 74:17–37, 1966.

[211] L. Garrote and R.L. Bras. A distributed model for real–time flood forecasting using digital elevation models. *Journal of Hydrology*, 167:279–306, 1995.

[212] M. Al-Smadi. *Incorporating Spatial and Temporal Variation of Watershed Response in a GIS–Based Hydrologic Model.* PhD thesis, Virginia Polytechnic Institute and State University, 1998.

[213] K.J. Beven and M.J. Kirkby. A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin*, 24(1):43–69, 1979.

[214] I.D Moore and J.P. Wilson. Length-slope factors for the revised universal soil loss equation: simplified method of estimation. *Journal of Soil and Water Conservation*, 47:423–428, 1992.

[215] G.R. P Foster. *Process-based modelling of soil erosion by water on agricultural land.* John Wiley and Sons, 1990.

[216] Branco Kaucic. Approximative terrain guarding with given number of guards. *Journal of Computing and Information Technology*, 13(4):307–312, 2005.

[217] R. Dikau. The application of a digital relief model to landform analysis in geomorphology. In J. Raper, editor, *Three Dimensional Applications in Geographical Information Systems*, pages 51–77. Taylor & Francis, London, 1989.

[218] E.E. Dikau, R. Brabb and R.M Mark. *Landform classification of New Mexico by computer.* U.S. Dept Interior, U.S. Geological Survey., 1991.

[219] E.H. Hammond. Small scale continental landform maps. *Annals of Association of American Geographers*, 44:32–42, 1954.

[220] B.J. Irvin, S.J. Ventura, and Slater B.K. Landform classification for soil-landscape studies. In *Proceedings of the Annual ESRI User Conference.* ESRI, 1995.

[221] Tomislav Hengl and David G. Rossiter. Supervised landform classification to enhance and replace photo-interpretation in semi-detailed soil survey. *Soil Science Society America Journal*, 67(5):1810–1822, 2003.

[222] P.A. Burrough, P.F.M. van Gaans, and R. A. MacMillan. High resolution landform classification using fuzzy-k-means. *Fuzzy Sets and Systems*, 113(1):37–52, 2000.

[223] T. Hengl, S. Gruber, and D.P. Shrestha. *Digital Terrain Analysis in ILWIS.* Lecture notes, International Institute for Geo-Information Science & Earth Observation (ITC), Enschede, 2003.

[224] J.A Richards. *Remote Sensing Digital Image Analysis: An Introduction.* Springer Verlag, 1986.

[225] S. Vassilopoulou, L. Hurni, V. Dietrich, Baltsavias, Pateraki E., M., E. Lagios, and I. Parcharidis. Orthophoto generation using ikonos imagery and high-resolution dem: a case study on volcanic hazard monitoring of nisyros island (greece). *ISPRS Journal of Photogrammetry and Remote Sensing*, 57:24–38, 2002.

[226] K.N. Liou. *An Introduction to Atmospheric Radiation*, volume 84 of *International Geophysics Series*. Academic Press, 2nd edition, 2002.

[227] Y. J. Kaufman. The atmospheric effects on remote sensing and its correction. In G. Asrar, editor, *Theory and Applications of Optical Remote Sensing*, pages 336–428. John Wiley & Sons, 1989.

[228] R. M. Haralick and L. G. Shapiro. *Computer and robot vision. Vol. I,, Reading, p.* Addison Wesley, 1992.

[229] H. Kishan Iddamsetty. Segmentation of range images for modelling of large outdoor scenes. Master's thesis, The University of Tennessee, Knoxville, 2003.

[230] Luong Chi Mai. *Introduction to computer vision and image processing*. Department of Pattern Recognition and Knowledge Engineering, Institute of Information Technology, Hanoi, Vietnam, 1998.

[231] R. C. Gonzales and R. E. Woods. *Digital Image Processing*. 1993.

[232] L. Lucchese and S.K. Mitra. A new method for color image equalization. In *Proc. of 2001 Int'l Conference on Image Processing (ICIP 2001), Thessaloniki, Greece, Sept. 2001, Vol. I*, 133-136.

[233] P. M. Mather. *Computer Processing of Remotely-sensed Images*. John Wiley & sons, 1999.

[234] W. J. Carper, T. M. Lillesand, and R.W. Kieffer. The use of intensity-hue-saturation transformations for merging spot panchromatic and multispectral image data. *Photogrammetric Engineering and Remote Sensing,*, 56:459–467, 1990.

[235] J.D. Foley, A. van Dam, S.K. Feiner, and J.F. Hughes. *Computer graphics: principles and practice*. Addison-Wesley Publishing Company, Inc, 1997.

[236] N.E.G. Roller and S. Cox. Comparison of landsat mss and merged mss/rbv data for analysis of natural vegetation. In *Proc. of the 14th International Symposium on Remote Sensing of Environment, San Jose, Costa Rica, 23-30 April*, pages 1001–1007, 1980.

[237] T.Y. Zhang and C.Y. Suen. A fast parallel algorithm for thinning digital patterns. *Communications of the ACM*, 27(3):236–329, 1984.

[238] J. G. Ruby and R. L. Fischer. Spectral signatures database for remote sensing applications. In S. S. Shen, editor, *Imaging Spectrometry VIII*, volume 4816, pages 156–163, 2002.

[239] M.A. Gilabert, J Gonzalez-Piqueras, and J. Garcia-haro. Acerca de los indices de vegetacion. *Revista de Teledeteccion*, 8, 1997.

[240] J. W. Rouse, R. H. Haas, J. A. Schell, and D. W. Deering. Monitoring vegetation systems in the great plains with erts. In *Proceding of the Third ERTS Symposium, NASA SP-351 I*, pages 309–317., 1973.

[241] J.M. Van der Knijff, R.J.A. Jones, and L. Montanarella. Soil erosion risk assessment in italy. Technical report, European Soil Bureau, 1999.

[242] B.N. Rock, D.L. Williams, and J.E. Vogelmann. Field and airborne spectral characterization of suspected acid deposition damage in red spruce (picea rubens) from

vermont. In *Proc., Symp. on Machine Processing of Remotely Sensed Data, Purdue Univ., West Lafayette, Indiana.* 1985.

[243] N.J. Lee and K. Nakane. Forest vegetation classification and biomass estimation based on landsat tm data in a mountainous region of west japan. *Forestry Sciences*, 50:159–171, 1997.

[244] A.J. Richardson and C.L. Wiegand. Distinguishing vegetation from background soil information. *Photogrammetric Engineering and Remote Sensing*, 43:1541–1552, 1977.

[245] C. Jr. Perry and L. F. Lautenschlager. Functional equivalence of spectral vegetation indices. *Remote Sensing and the Environment*, 14:169–182, 1984.

[246] A. Bannari, A. R. Huete, D. Morin, and Zagolski. Effets de la couleur et de la brillance du sol sur les indices de végétation. *International Journal of Remote Sensing*, 17(10):1885–1906, 1996.

[247] J. Qi, A. R. Chehbouni A. andHuete, Y. H. Kerr, and S. Sorooshian. A modified soil adjusted vegetation index. *Remote Sensing and the Environment*, 48:119–126, 1994.

[248] A. R. Huete. A soil-adjusted vegetation index (savi). *Remote Sensing and the Environment*, 25:53–70., 1988.

[249] R.J. Kauth and G.S. Thomas. The tasseled cap - a graphic description of the spectral temporal development of agricultural crops as seen by landsat. In *Proceedings of the Symposium on Machine Processing of Remotely Sensed Data. Perdue University, West Lafayette, Indiana*, pages 41–51. 1976.

[250] Brian Price. Image vectorization. Master's thesis, Department of Computer Science, Brigham Young University, 2006.

[251] K. Ananthanarayanan and K. Varghese. Vectorization of contours from scanned topographical map. *GIS Development. Asia Pacific Edition.*, 10, 2006.

[252] J. Pouderoux and S. Spinello. Contour line recognition from scanned topographic maps. In *9th International Conference on Document Analysis and Recognition*, pages 779–783, 2007.

[253] P. Arrighi and P. Soille. From scanned topographic maps to digital elevation models. In *Proceedings of Geovision 99*, 1999.

[254] Y. Chen, R. Wang, and J. Qian. Extracting contour lines from common-conditioned topographic maps. *IEEE Transactions on Geoscience and Remote Sensing*, 44(4):1048–1057, 2006.

[255] F. Leberl and D. Olson. Raster scanning for operational digitizing of graphical data. *Photogrammetric Engineering and Remote Sensing*, 48(4):615–627, 1982.

[256] D. Greenlee. Raster and vector processing for scanned line work. *Photogrammetric Engineering and Remote Sensing*, 53(10):1383–1387, 1987.

[257] T. Pavlidis. *Algorithms for Graphics and Image Processing*. 1982.

[258] P. Brunet and LL. Pérez Vidal. Smooth contour line construction with spline interpolation. *Questiio*, 8(3):133–144, 1984.

[259] M.A. Sabin. Contouring. a review of methods for scattered data. In *Mathematical*

methods in computer graphics and design. Academic Press, 1980.

[260] D. C. Sutcliffe. Contouring over rectangular and skewed rectangular grids. an introduction. In *Mathematical methods in computer graphics and design*. Academic Press, 1980.

[261] W. V. Snyder. Contour plotting. *ACM Transactions on Mathematical Software*, 4(3):290–294, 1978.

[262] W. Lorensen and H. E. Cline. Marching cubes: A high resolution 3d surface construction algorithm. In *Computer Graphics (SIGGRAPH 87 Proceedings)*, volume 21, pages 163–170, 1987.

[263] Z. Chen and J.A. Guevara. Systematic selection of very important points (vip) from digital terrain models for construction triangular irregular networks. In *Proceedings AutoCarto 8*, pages 50–56, 1987.

[264] R.J. Fowler and J.J. Little. Automatic extraction of irregular network digital terrain models. *Computer Graphics*, 13:199–207, 1979.

[265] Jay Lee. A drop heuristic conversion method for extracting irregular network for digital elevation models. In *GIS/LIS '89 Proc.*, volume 1, pages 30–39. American Congress on Surveying and Mapping, Nov. 1989.

[266] N. Dyn, D. Levin, and S. Rippa. Data dependent triangulation for piecewise interpolation. *IMA Journal of Numerical Analysis*, 10(1):137–154, 1990.

[267] B. Delaunay. Sur la sphère vide. *Otdelenie Matematicheskikh i Estestvennykh Nauk*, 8:793–800, 1934.

[268] Peter Su and R. L. Scot Drysdale. A comparison of sequential delaunay triangulation algorithms. In Association for Computing Machinery, editor, *Proceedings of the Eleventh Annual Symposium on Computational Geometry*, pages 61–70, 1995.

[269] Steven Fortune. *Computing in Euclidean Geometry*, chapter Voronoi Diagrams and Delaunay Triangulations, pages 193–233. Lecture Notes Series on Computing. World Scientific, 1992.

[270] D. Andrews. Simplifying terrain models and measuring terrain model accuracy. Master's thesis, Dept. of CS, UBC, 1996.

[271] J. O'Rourke. *Computational Geometry in C*. Cambridge University Press, 1998.

[272] N. Megiddo. Linear-time algorithms for linear programming in r3 and related problems. *SIAM Journal on Computing*, 12(4), 1983.

[273] Skyum. A simple algorithm for computing the smallest enclosing circle. *Information Processing Letters*, 3:121–125, 1991.

[274] D.H. Douglas and T.K. Peucker. Algorithms for the reduction of the number of points required to represent a line or its character. *The American Cartographer*, 10(2):112–123, 1973.

[275] R.B. McMaster. Automated line generalization. *Cartographica*, 24(2):74–111, 1987.

[276] R.B. McMaster. The integration of simplification and smoothing algorithms in line generalization. *Cartographica*, 26(1):101–121, 1989.

[277] D. H. Douglas. Least-cost path in gis using an accumulated cost surface and slope lines. *Cartographica*, 31:37–51., 1994.

[278] A. Gatrell. *Distance and space: a geographical perspective.* Clarendon Press, 1983.

[279] J.R. Eastman. Pushbroom algorithms for calculating distances in raster grids. In *Proceedings, AutoCarto 9*, 1989.

[280] Chaoqing Yu. Extensions to least-cost path algorithms for roadway planning. *International Journal of Geographical Information Systems*, 17(4):361–376, 2003.

[281] E. Langmuir. *Mountaincraft and leadership.* The Scottish Sports Council/MLTB., Cordee, Leicester, 1984.

[282] E. W. Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1:269–271, 1959.

[283] G. Gallo and S. Pallottino. Shortest paths algorithms. *Annals of Operations Research*, 13:3–79, 1988.

[284] R. C. Prim. Shortest connection networks and some generalisation. *Bell System Technical Journal*, 36:1389–1401, 1957.

[285] Joseph. B. Kruskal. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society*, 7(1):48–50, 1956.

[286] Jaroslav Nesetril, Eva Milkova, and Helena Nesetrilova. Otakar boruvka on minimum spanning tree problem (translation of both 1926 papers, comments, history). *DMATH: Discrete Mathematics*, 233, 2001.

[287] D. Cheriton and R. E. Tarjan. Finding minimum spanning trees. *SIAM J. Computing*, 5:724–742, 1976.

[288] K.R. Gabriel and R.R. Sokal. A new statistical approach to geographic variation analysis. *Systematic Zoology*, 18:259–278, 1969.

[289] P. Swain and S. Davis. *Remote Sensing: The Quantitative Approach.* Mc Graw-Hill, 1978.

[290] J.A. Bendiktsson, P.H. Swain, and O.K. Ersoy. Neural network approaches versus statistical methods on classification of multisource remote sensing data. *IEEE Trans. Geosci. Remote Sensing*, 28(4):540–552, 1990.

[291] Edward H. Bosch. Texture information and supervised classification of hyperspectral imagery by means of neural networks. In *Proceedings GeoComputation 1999*, 1999.

[292] F. Hepner. Artificial neural network classification using a minimal training set: Comparison to conventional supervised classification. *Photogrammetric Engineering and Remote Sensing*, 56(4):469–473, 1990.

[293] J.D. Paola and R.A. Schowengerdt. A detailed comparison of backpropagation neural network and maximum-likelihood classifiers for urban land use classification. *IEEE Transactions on Geoscience and Remote Sensing*, GE-33(4):981–996, 1995.

[294] G. M. Foody. Approaches for the production and evaluation of fuzzy land cover classifications from remotely sensed data. *Int. J. Remote Sensing*, 17(7):1317–1340, 1996.

[295] E. Forgy. Cluster analysis of multivariate data: Efficiency vs. interpretabilityof classifications. *Biometrics*, 21:768, 1965.

[296] T. A. Roth, G. Strunz, and S. Dech. Object-oriented classification of landsat-7 data for regional planning purpose. *The international archives of the photogrammetry, remote sensing and spatial information sciences*, 34(7), 2003.

[297] G. Castilla. *Object-oriented analysis of remote sensing images for land cover mapping: conceptual foundations and a segmentation method to derive a baseline partition for classification.* PhD thesis, Universidad Politécnica de Madrid. Escuela Superior de Ingenieros de Montes, 2003.

[298] J. Anderson, E. Hardy, Roach. J., and R. Witmer. A land use and land cover classification system for use with remote sensor data. Professional Paper 964, U.S. Geological Survey, 1976.

[299] M. F. Goodchild. Integrating gis and remote sensing for vegetation analysis and modelling: methodological issues. *J. Veg. Sci.*, 5:615–26, 1994.

[300] L. Plourde and R. Congalton. Sampling method and sample placement: How do they afect the accuracy of remotely sensed maps? *Photogramm. Eng. Rem.*, 69(3):289–297, 2003.

[301] J.R. Jensen. *Introductory digital image processing: a remote sensing perspective.* Series in Geographic Information Science. Prentice Hall, 1996.

[302] P M. Mather. Theoretical problems in image classification. *Proceedings of the Royal Society*, 326:127–135, 1990.

[303] I. L. Thomas, V.M. Benning, and N.P. Ching. *Classification of Remotely Sensed Images.* Adam Hilger, 1987.

[304] A.S. Fotheringham, C. Brunsdon, and M.E. Charlton. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships.* Wiley, 2002.

[305] K. Tanaka. *An Introduction to Fuzzy Logic for Practical Applications. Springer.* Springer Verlag, 1997.

[306] H.J. Zimmermann. *Fuzzy Set Theory and Its Applications, Fourth Edition. Kluwer.* Kluwer Academic Publishers, 2001.

[307] D.J. Maguire and M. Batty, Goodchild. *GIS, spatial analysis and modeling.* ESRI Press, 2005.

[308] B. M. Ayyub. *Elicitation of Expert Opinions for Uncertainty and Risks,.* CRC Press, 2001.

[309] G. Shafer. *A Mathematical Theory of Evidence.* Princeton University Press, 1976.

[310] J. Gordon and E.H. Shortliffe. A method for managing evidential reasoning in a hierarchical hypothesis space. *Artificial Intelligence*, 26:323–357, 1985.

[311] Karl Sentz and Scott Ferson. Combination of evidence in dempster-shafer theory. Technical report, Sandia National Laboratories, 2002.

[312] T.L. Saaty. A scaling method for priorities in hierarchical structures. *J. Math. Psychology*, 15:234–281, 1977.

[313] T. Fung and E. LeDrew. Application of principal component analysis to change detection. *Photogrammetric Engineering and Remote Sensing*, 53(12):1649–1658, 1987.

[314] G. F. Byne, P. F. Crapper, and K. K. Mayo. Monitoring land-cover change by principal component analysis of multitemporal landsat data. *Remote Sensing of Environment*, 10:175–184, 1980.

[315] M. J. Kraak and A. Brown, editors. *Web cartography*. Francis and Taylor, 2001.

[316] R.L.G. Lemmens, A. Wytzisk, and R.A. de By. Integrating semantic and syntactic descriptions to chain geographic services. *IEEE internet computing*, 10(5):42.52, 2006.

[317] ESRI. Gis best practices. mobile gis, 2007.

[318] A. J. Brimicombe. Gis - where are the frontiers now? In *Proceeding GIS 2002*, pages 33–45, 2002.

[319] A. A. Solyman. Investigating mobile gis. *Directions Magazine*, November, 2005.

[320] X. Lopez. *Telegeoinformatics*, chapter Location-Based Services, pages 171–188. CRC Press, 2004.

[321] A. Dey. Understanding and using context. *Personal and Ubiquitous Computing*, 5(1):4–7, 2001.

[322] B. Schilit, N. Adams, and R. Want. Context-aware computing applications. In *Proceedings of IEEE Workshop on Mobile Computing Systems and Applications*, pages 85–90, 1994.

[323] A-M. Nivala and L.T. Sarjakoski. An approach to intelligent maps: Context awareness. In *MobileHCI'03, 5th International Conference on Human Computer Interaction with Mobile Devices and Services, September 8th, 2003, Udine, Italy*, 2003.

[324] T. Reichenbacher. *Mobile Cartography - Adaptive Visualisation of Geographic Information on Mobile Devices*. PhD thesis, 2004.

[325] B. H. McCormick. Visualization in scientific computing. *Computer Graphics*, 21(6), 1987.

[326] T. M. Rhyne. Going virtual with geographic information and scientific visualization. *Computers and Geosciences*, 23(4):489–491, 1997.

[327] A.M. MacEachren. *Visualizacion in modern cartography*, chapter Visualization in modern cartography: setting the agenda, pages 1–12. Pergamon Press, Inc., 1994.

[328] J. Bertin. *Cartes et figures de la Terre*. Ed. G. Pompidou, 1987.

[329] A.M. MacEachren. *How Maps Work: Visualization, Representation and Design*. Guilidford Press.

[330] A. Treisman and S. Gormican. Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, 95:15–48., 1988.

[331] G. F. Jenks. The data model concept in statistical mapping. *International Yearbook of Cartography*, 7:186–190, 1967.

[332] R. Arnheim. *El pensamiento visual*. Paidos, 1986.

[333] J.L. Morrison. *The computer in contemporary cartography*, chapter Computer technology and cartographic change, pages 5–23. John Wiley and Sons, 1980.

[334] C. Haeberling. Cartographic design principles for 3d maps - a contribution to cartographic theory.

[335] A.M. MacEachren. *Visualization in Geographical Information Systems*, chapter Time as a cartographic variable, pages 115–130. John Wiley, 1994.